THREE ESSAYS ON INFRASTRUCTURE INVESTMENT IN CHINA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yu Qin

August 2014

THREE ESSAYS ON INFRASTRUCTURE INVESTMENT IN CHINA

Yu Qin, Ph.D.

Cornell University 2014

My dissertation examines the consequences of the massive infrastructure investment in China from three different aspects. In the first chapter, I examine the distributional impacts of high-speed rail upgrade in China. By exploiting the quasi-experimental variation in whether counties were affected by this project, I find that relative to non-affected counties, affected counties on the upgraded railway lines suffered a 4-6 percent reduction in GDP and GDP per capita following the upgrade, which can be explained largely by the concurrent drop in fixed asset investment. In addition, since high-speed rail upgrade affects transportation of passengers and not transportation of goods, the negative impact of the upgrade is significant in the service sector and not in the manufacturing sector. A possible mechanism consistent with the core-periphery model is also discussed in this chapter. In the second chapter, I use four waves of a primary panel household survey conducted in 17 remote natural villages in China to study how road access shapes farmers' production patterns, input use, and rural poverty. The results show that access to roads facilitates specialization in agricultural production. In natural villages with better road access, farmers plant fewer numbers of crops, purchase more fertilizer, and hire more labor. Consequently, road connections improve household agricultural income and reduce poverty. However, better access to rural roads does not appear to bring about significant changes in nonagricultural income. In the third chapter, I study whether the increasing insecurity of home ownership being reported in

the media induces urban households to save more in Chinese cities, which is in favor of the precautionary savings motive. Using Difference GMM models, I find that worse insecurity of home ownership, as indicated by more frequent forced evictions, leads to higher household savings rate at the prefecture city level. In addition, I find that the impact could work through a reduction of home purchase due to forced evictions, as well as precautionary savings motive of existing home owners.

BIOGRAPHICAL SKETCH

Yu Qin, born in July, 1987, is a PhD candidate in the Dyson School of Applied Economics and Management at Cornell University. Her fields of interests include development economics and urban economics. Before coming to Cornell University in 2009, she graduated from Peking University in China with a Bachelor of Arts in Finance and a Bachelor of Science in Statistics.

This document is dedicated to my husband Jijie Zhao, without whom I cannot survive the winters in Ithaca.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my PhD committee: Professor Ravi Kanbur, Professor Nancy Chau, Professor Damon Clark and Professor Xiaobo Zhang for their guidance and continuous support during my PhD study. Their wisdom and passion on research inspire me to continue my journey in the academic world.

I am also indebted to my parents, Yuehong Qin and Changyun Sun, and my husband, Jijie Zhao, for their valuable support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Massive infrastructure investment is iconic to China's economic development in the past decade. My dissertation tries to understand the microeconomic and macroeconomic consequences of infrastructure investment in China.

In Chapter 2, I examine the distributional consequences of high-speed rail upgrade in China. As an urban-biased transportation technology, high-speed rail connects large cities with bullet trains. However, such high-speed trains do not stop in rural counties in order to maintain the high speed. Even worse, the bullet trains squeeze out some conventional train services since they share the same railroad with the slow trains after the railway upgrade. In this chapter, I exploit the quasi-experimental variation in whether counties were affected to examine the distributional impacts of the upgrade. More specifically, I construct instruments for whether a county is affected and compare affected and non-affected counties before and after the high-speed upgrade. The main results suggest that counties being affected by high-speed rail upgrade experienced 4-6 percent GDP and GDP per capita reduction, which can explain around 64 percent of the predicted GDP growth differentials between the affected and non-affected counties. The reduction in GDP per capita is not driven by population changes. Instead, I find that the magnitude of the reduction in fixed asset investment almost explains the reduction in GDP. Intuitively, when the cities had been more conveniently connected by the bullet trains, investment left the counties and crowded into the cities in pursuit of higher returns. In addition, I find that the negative impact of the upgrade is significant in the service sector and not in the manufacturing sector. This is reasonable since high-speed rail upgrade affects the passenger rail service instead of the freight rail service. I also discuss the potential mechanisms of such impact. These findings imply that the distributional implications of these types of investments can be dramatic, suggesting that investment policies might have to be augmented by supplementary policies designed to mitigate these distributional consequences.

In Chapter 3, I try to understand how rural road access shapes farmers' production patterns, input use, and rural poverty. I use four waves of a primary panel household survey conducted in 17 remote natural villages in China to examine this question. The results suggest that access to roads facilitates specialization in agricultural production. In natural villages with better road access, farmers plant fewer numbers of crops, purchase more fertilizer, and hire more labor. Consequently, road connections improve household agricultural income and reduce poverty. The first two chapters provide important policy implications for China. In the past several decades, the Chinese government has made significant investments in building nationwide highway and high-speed rail systems. As the highway and high-speed rail density increases, the marginal returns to such investments are likely to decrease. In addition to that, the distributional consequences of highway and high-speed rail investment may exacerbate the urban-rural disparities in China. Therefore, it may make more economic sense to gear investment toward local road construction (such as rural road) in areas lack of transportation infrastructure.

In Chapter 4, I study the macroeconomic consequences of a rising social problem due to infrastructure investment in China: forced evictions. As local governments have been aggressively investing in infrastructure, the residential areas targeted by such investments need to be demolished. Therefore, the incidences of forced evictions have been increasing dramatically both at the intensive (more cases in the same area) and extensive (more areas being affected) margin in recent years. The number of news about forced evictions by searching Baidu News Archive, the world's largest Chinese news search engine, rose from 307 in 2004 to 125,000 in 2012. However, there is no paper studying the potential consequences of such rising social problem in China as far as is concerned. Previous literature suggests that home ownership in China is associated with reduced household savings (Chamon and Prasad, 2010). It is thus interesting to investigate whether owning a home with expropriation uncertainty due to forced evictions will induce households to save more. I use the annual incidence of forced evictions reported in the news at the prefecture city level to measure the frequency of forced eviction from 2004 to 2011, which is collected from Baidu News Archive. The household savings rate is collected from the CEIC database for the same time period. By employing the dynamic panel data model, I find that worse insecurity of home ownership, as indicated by more frequent forced evictions, leads to higher household savings rate at the prefecture city level. In addition, I find that the impact could work directly through a reduction of home purchase due to forced evictions, or indirectly through precautionary savings behavior of existing home owners.

CHAPTER 2

"NO COUNTY LEFT BEHIND?": THE DISTRIBUTIONAL IMPACT OF HIGH-SPEED RAIL UPGRADE IN CHINA

2.1 Introduction

Infrastructure investments are regarded as key instruments to promote overall economic growth. However, such investments are not evenly distributed across different regions of a country, possibly due to differences in expected returns, budget constraints, planning concerns, and so on. For example, in the most recent highway construction boom in China, the highway length per capita in the affluent Guangdong Province quadrupled from 2003 to 2010, while it increased by only one fourth in the relatively underdeveloped Guizhou Province during the same period.[1] Therefore, the regions or sectors more "local" to investments (such as Guangdong Province) may benefit more than less-affected regions or sectors (such as Guizhou Province). The distributional consequences will be even more pronounced if investments biased toward one sector or region hurt less-affected sectors or regions.

In this chapter, I explore this possibility by investigating the distributional impacts of one such infrastructure investment: high-speed rail upgrade in China. There are several reasons why this is a useful case to study. First, investment in high-speed rail is prevalent in both developed and developing countries. Currently, more than 10 countries in the world have high-speed rail, including Spain, Japan, Germany, France, China, the United States, Belgium, Italy, the Netherlands, the United Kingdom, Korea, Taiwan (China), and Turkey (UIC,

[1] China Statistical Yearbooks, 2004-2011.

2010). As well, a number of other countries, such as India, Russia, Brazil and Canada, plan to upgrade their railway lines into high-speed rail (UIC, 2010). China is a very relevant country to study high-speed rail's impact since such investment in the country is very large in scale. Among all the infrastructure investment in China, investment in railroads takes up to 23 percent with a high growth rate in recent years.[2] The annual investment in the railroad sector in year 2010 (around 120 billion US Dollars) was more than ten times that of the investment in 2003 (around 10 billion US Dollars). A large proportion of investment in the railroad sector was for high-speed rail upgrades on existing railway lines and the construction of new high-speed rail. After the high-speed rail upgrade in 2007, the total length of high-speed rail reached 6,000 kilometers in China, top in the world even today.

Second, like all such investments, high-speed rail upgrades in China are known to favor urban areas. High-speed rail, by definition, is a type of passenger rail transport that travels at speeds above 200 kilometers per hour. In order to maintain the high speed, the featured service of high-speed rail, the bullet trains, stop only in populous urban areas, where there are higher demands for time savings, in contrast with small cities and rural areas. Thus, counties with upgraded railway lines may find bullet trains bypassing them (Economist, 2011). Indeed, as suggested by Figure A.1, around 3,000 out of around 6,100 passenger train stops in China have been abandoned in the past ten years due to the speed acceleration of passenger train services, especially after year 2004, when high-speed rail upgrading began. That is to say, even though the high-speed trains help facilitate economic activities across cities due to significantly less travel time, they may actually hurt the small counties along the acceler-

[2] China Statistical Yearbooks, 2004-2011.

ated railway lines by passing them by and depriving them of access. In the urban planning literature, this is known as the "tunnel effect," defined as "an improvement in access to major cities but at the expense of breaking up the space between them. The increase in dynamism in large nodes is compensated by a decrease in the activity of areas between the connection points" (Albalate and Bel, 2012). The latest World Bank report on China's high-speed rail development also documents the fact that some conventional train services were removed after the introduction of high-speed rail (Bullock et al., 2012).

Third, whether the non-targeted counties have been affected by the upgrade process has been quasi-random to a large extent, which facilitates credible empirical analysis on the causal impact of high-speed rail upgrade on such affected counties. The two rounds of high-speed rail upgrade, parts of China's railway speed acceleration project since 1997, were implemented in the year 2004 and 2007. There are two reasons why the upgrade has been quasi-random for the affected counties. First, all the upgrades were implemented on existing railway lines, which mitigates the concern of the selection problem on high-speed rail placement. Second, as the selection for high-speed rail upgrade mainly depends on which large cities the existing railways are connected to, the counties in between cities affected by the speed acceleration can be regarded as quasi-random since they were not selected on purpose (Michaels, 2008; Datta, 2012).

I exploit this quasi-experimental variation in whether counties were affected to examine the distributional impacts of the upgrade. Specifically, I examine the impact of high-speed rail upgrade by comparing the economic outcomes of the counties located on the affected railway lines with the counties located on non-affected railway lines, before and after, using county level statistics collected from statistical yearbooks and other published statistical reports. I first apply a difference-in-difference setting in order to compare the high-speed rail affected counties and non-affected counties, before and after. The common trend assumption required by difference-in-difference analysis satisfies since the pre-trends of outcome variables are similar between control and treatment groups. To strengthen my estimation, I also instrument the selection of affected counties by whether a county is located on the five main railway lines in China (four lines connecting Beijing to the north: Haerbin; south: Guangzhou and Hong Kong; east: Shanghai, while one connects Lianyungang in the east to Urumqi in the west). The assumptions for the instrumental variable estimation are that counties located on the main railway lines are associated with a higher likelihood of being selected into high-speed rail upgrades, while their placement on main railway lines affects economic development only through the impact on speed acceleration. Even if counties located along the main railway lines are selected due to their greater economic potential, it will only bias my estimate downward as I expect the counties located on the accelerated railway lines are negatively affected in the later period of the project.

My analysis conveys several main findings. First, the estimations using OLS and Two-Stage Least Square (2SLS) consistently suggest that being located on the high-speed railway lines decreases a county's total GDP and GDP per capita by 4-6 percent on average, which is around 336-503 million *yuan* (54-81 million US Dollars), given the average county level GDP as 8.39 billion *yuan* (around 1.35 billion US dollars) in 2006 in the affected regions. In addition, the negative impact due to high-speed rail upgrade can explain around 64% of the predicted GDP growth rate differentials right after the upgrade between the affected and non-affected counties. The results still hold if I collapse the panel data from multiple years into two periods, i.e., “pre” and “post,” following the suggestion of Bertrand et al. (2004) on correctly estimating the standard errors in difference-in-difference analysis. Such implies that the urban-biased investment of high-speed rail upgrades hurt the economic growth of non-targeted counties located on the upgraded railways.

Second, the reduction of GDP is likely to be investment driven, as evidenced by the 10-11 percent reduction of fixed asset investment in the affected counties. This is not surprising since investment is a driving force of GDP growth in China (Qin et al. 2006; Yu, 1998). Intuitively, when the cities had been more conveniently connected by high-speed trains, investment left the counties and crowded into the cities in pursuit of higher returns due to expected growth. The result still holds for robustness checks.

Third, the impact of the railway upgrade varies in different sectors. Since high-speed rail significantly reduces transportation cost of passengers rather than goods, its negative impact in the affected counties is more pronounced in the service sector than in the manufacturing sector. More precisely, the growth rate of service sector value added is reduced by 3-4 percent after the high-speed rail upgrade, while the growth rate of industrial sector value added is not significantly affected.

Fourth, I discuss the channels that may account for the investment-driven economic slowdown in the affected counties. Specifically, I test two possible channels: 1) increases in trade cost due to reduced train services in the affected counties may lead to decreases in economic activities; 2) increases in agglomeration spillovers with a more tightly connected transportation network between large cities may divert economic activities from counties to populous urban districts. I find that the second channel plays a more important role in explaining the negative impact of high-speed rail upgrade.

The contributions of this chapter are threefold. First, to my knowledge, this is the first paper documenting the distributional consequences of high-speed rail projects to the non-targeted rural areas, which complements the rich body of literature examining the causal relationship between access to infrastructure and various aspects of economic development in both developing and developed countries (Ahlfeldt, 2011; Atack et al. 2010; Banerjee et al. 2009, 2012; Baum-Snow 2007; Baum-Snow et al. 2012; Donaldson 2013; Donaldson and Hornbeck 2013; Duflo and Pande 2007; Faber 2013; Zheng and Kahn, 2013). Specifically, this paper provides the first set of results demonstrating an investment-driven GDP reduction in the non-targeted rural areas affected by infrastructure investment. Second, the main findings of this paper provide an empirical test to the core-periphery model (Fujita et al. 2001), especially its recent development by introducing the service sector (Leite et al. 2013). The evidence in this paper suggests that the periphery rural areas may experience a reduction in service sector output when transportation cost decreases. Third, this paper also provides useful insights in understanding the increasing rural-urban disparity in China in the past few decades, where urban biased infrastructure investment may have played a role (Kanbur and Zhang 2005; Xu 2011).

The chapter is organized as follows: Section 2.2 describes the policy background of high-speed rail upgrade in China. Section 2.3 describes the identification strategy and data sources. Section 2.4 shows the main findings and robustness checks. Section 2.5 discusses the heterogeneous impacts of the railway upgrade in different sectors, possible channels the impact may work through and the magnitude of such impact. Finally, Section 2.6 concludes.

2.2 Background of China's High-Speed Rail Upgrade

2.2.1 Railway network in China

China is the third Asian country to adopt a railroad system, after Japan and India. The first railroad in China, constructed in the year 1876 by the British, was a local railway near Shanghai. During the 73 years after the first railroad in China and before the founding of the People's Republic of China, around 23,000 kilometers of railroad were constructed in China. However, half of them were destroyed during World War II.

In 1949, railroad construction resumed and has been emphasized in almost all of China's "Five-Year Plans." By the late 1990s, the operating railroad length had been increased to around 66,000 kilometers, with six main railway lines connecting several largest cities in different parts of the country: 1) Beijing-Shanghai (*jinghu xian*); 2) Beijing-Haerbin (*jingha xian*); 3) Beijing-Guangzhou (*jingguang xian*); 4) Beijing-Hong Kong (*jingjiu xian*); 5) Lianyunguang-Urumqi (*longhai-lanxin xian*); 6) Beijing-Baotou (*jingbao xian*).

In late 2002, the new Minister of Railways, Zhijun Liu, proposed his "Great Leap Forward" strategy, which encouraged further expansion of the railroad network and many technology upgrades, including high-speed rail upgrades and construction (Liu, 2003). The *Mid-long Term Railway Network Plan* enacted by the State Council in 2005 set the goal of expanding railroad length to 100,000 kilometers by the end of 2020, which was further revised to 120,000 kilometers in the year of 2008, with a budget of around 4,000 billion *yuan* (State Council, 2004, 2008). By the end of 2007, all the provinces in China had been connected with railroad networks, as suggested in Figure A.2. However, it is clearly shown that the railroad coverage in the west, the relatively poor area, is significantly lower than in the east.

2.2.2 Railway speed acceleration and high-speed rail upgrade

Mainly in response to the profit loss under the competition of road and air transportation, China's Railways Ministry started several rounds of speed acceleration on existing railway lines spanning from 1997 to 2007. The project had two stages. In the first stage, train speed was increased gradually in the first four waves, namely 1997, 1998, 2000 and 2001. In 1997, the first round of speed acceleration was initiated on three main railway lines connecting from Beijing to Shanghai, Guangzhou, and Haerbin. The average passenger train speed was increased from around 48.1 kilometers per hour to 54.9 kilometers per hour. Subsequently in 1998, 2000 and 2001, another three waves of speed acceleration were implemented on the main railway lines, increasing the average train speed nationwide to 61.6 kilometer per hour by the end of 2001.

In the second stage, speed acceleration was targeted towards upgrading the existing railway into high-speed rail, with sustained speed greater than 200 kilometers per hour or higher. In 2004, around 1,960 kilometers of railroad had been upgraded to high-speed rail, with 19 pairs of city-to-city nonstop passenger trains operating on it. In 2007, the upgraded high-speed rail was expanded to around 6,000 kilometers with 257 pairs of China Railway High-speed (CRH) trains operating on a daily basis, which significantly shortened the commuting time between large cities. For example, the travelling time from Beijing to Fuzhou, the provincial capital of Fujian in the south of China, was reduced from around 33 hours to 19.5 hours with the introduction of CRH trains in 2007. The travelling time by train was reduced by more than half from Shanghai to Nanchang and Changsha, which are the two provincial capitals in southeast China. According to the vice Minister of the Chinese Railways Ministry, the travelling time between cities by CRH trains was reduced by an average of 20-30 percent (Sina 2007).

The dramatic expansion of high-speed rail in the year 2007 reflects the "Great Leap Forward" strategy proposed by the ex-Minister of the Chinese Railways Ministry, Zhijun Liu, who was removed due to corruption allegations in early 2011. During Liu's tenure, China invested a huge amount of money into railway expansion, upgrades, and construction of high-speed railway lines. As most of the high-speed railway lines were updated from existing railways, some slow train services on the upgraded lines were canceled in order to accommodate CRH trains. As a consequence, the number of operating slow trains significantly decreased with the increase of high-speed rail mileage. For example, in 2002—before high-speed rail upgrade—352 pairs of daily slow passenger trains operated nationwide. The number dropped to 224 in 2007.[3]

Even though high-speed rail benefits transportation from city to city, it may do harm to the economy of the small counties along the upgraded high-speed railway lines through two possible channels. First, the affected counties lose

[3]There is no significant difference in terms of capacity between high-speed rail passenger trains and normal passenger trains. A typical passenger train contains 16-20 coaches, with a capacity of 110 passengers in each coach.

their geographic advantage in the transportation network, leading to increases of transportation cost. Specifically, counties located on the high-speed railway lines experienced a reduction of slow train services on the upgraded high-speed rail route. Even though the total number of train services might not have decreased if the affected counties utilized other conventional railway lines as well, the number of train services still decreased in relative terms compared to the counterfactual without high-speed rail upgrade; i.e., some of the slow train services on the affected railway lines were not canceled. Therefore, due to increased transportation cost, the affected counties become less integrated, and economic activities decline in response to the increase in trade barriers.

Second, since the high-speed rail connects large cities more tightly, the economic activities in cities may generate more agglomeration externalities. Therefore, some economic activities in small counties may be diverted to large cities in order to enjoy the agglomeration spillovers. Both the above-mentioned channels lead to lower economic returns to railroads in those places, which affects the overall economic performance.

2.2.3 Program placement

In this chapter, I focus on the high-speed rail upgrade in 2004 and 2007.[4] As upgrading existing railway lines for speed acceleration is costly, not all the railway

[4] As mentioned in 2.2, there were four rounds of speed acceleration in 1997-2001 before the high-speed upgrade. We will not focus on that since the scale of the project is small compared to the 2004 and 2007 high-speed upgrade. None of the railway lines in China had been upgraded to high-speed rail before 2004. An impact evaluation on the speed acceleration in 1997-2001 using difference-in-difference is shown in Table B.12, which suggests little impact of the four rounds of speed-up on economic performance in the affected counties. However, in order to ensure a cleaner identification, I exclude the observations from 1997 to 2001 in the control group when estimating the impact of high-speed rail upgrade in 2004 and 2007.

lines were selected for upgrade. In 2004, the three main railway lines connecting Beijing to Haerbin, Shanghai, and Guangzhou were partially upgraded to high-speed rail, with around 20 pairs of nonstop bullet trains operating on them (Figure A.3). Later in 2007, the upgrading was completed on the above-mentioned three railway lines and on two additional main lines (Lianyungang to Urumqi and Beijing to Hong Kong,) as well as four other regional lines (Hangzhou to Zhuzhou, Guangzhou to Shenzhen, Wuhan to Jiujiang, and Qingdao to Jinan, Figure A.4).

In general, the priority of high-speed rail upgrade was given to the main railway lines first, as they connect big cities, such as Beijing, Shanghai, and Guangzhou, which generate huge demand for railway transportation. Besides the main lines, several regional railway lines were selected for upgrade, as they pass through regions of high economic growth, such as the Pearl River Delta Region (Guangzhou to Shenzhen) and the affluent regions in Zhejiang Province (Hangzhou to Zhuzhou). I will argue in the next section that while these placement patterns were not random, they still facilitate credible identification strategies.

2.3 Data and identification

2.3.1 Identification strategy

The goal of this chapter is to reveal the distributional impact of high-speed rail upgrade in China. Specifically, the urban biased high-speed rail upgrade may hurt the economic growth of non-targeted counties/regions when it improves the connection between urban areas. In order to test the above hypothesis, the difference-in-difference strategy is applied to compare the counties located on the affected railway lines to the counties located on other railway lines, before and after each round of high-speed rail upgrade. It is worth emphasizing that all the urban districts in prefecture level cities have been excluded from the sample since they are likely to be selected on purpose in the high-speed rail upgrade projects.

The key assumption in difference-in-difference analysis is common trend. In this case, it would be violated if counties in the control group and treatment group have different growth patterns prior to high-speed rail upgrade. To test the common trend assumption, I use an event study analysis to show that the control group and treatment group have similar pre-trend in terms of GDP, per capita GDP and fixed asset investment before the upgrade process. More details about the event study are discussed in Section 2.4.3.

A problem posed by difference-in-difference analysis is the non-random placement of the treatment group. That is, in our context, the placement of high-speed rail upgrade is not randomly selected. However, the quasi-experimental nature of high-speed rail upgrade at the county level renders the non-random placement problem much less a concern for two reasons. First, all the upgrades were implemented on existing railway lines, which mitigates some of the concerns in the selection problem of high-speed rail placement.[5] Second, as the selection of affected railway lines mainly depends on the cities it connects, rather

[5]In addition to high-speed rail upgrade, China also constructed new high-speed rails, such as high-speed rail from Beijing to Shanghai and Wuhan to Guangzhou. In observance to the fact that the earliest new high-speed rails started to operate in December 2009, I exclude the county statistics after year 2009 in the estimation to avoid the possible intertwined impact of new high-speed rails and high-speed rail upgrade due to network effect. In addition, the counties being affected by newly constructed high-speed rails are also excluded from the estimation.

than the counties it bypasses, it can be treated as a quasi-natural experiment for the counties located on railway lines. This argument is similar to that of Michaels (2008) and Datta (2012), both of whom argue that if a highway is built to connect two cities, it must pass through areas that lie between the two, which affects the outcomes in such areas as a quasi-random shock. However, for regional lines which pass through only a few counties and cities, the decision to select the railway lines for high-speed rail upgrade may also depend on the counties they cross. Such counties may in fact bias my estimation.

Therefore, I employ an instrumental variable to identify the program selection. Specifically, I use whether a county is located on the main railway lines to identify whether it is affected by the speed acceleration or not. As mentioned in Section 2.2.1, there are six main railway lines in China. I exclude the railway line from Beijing to Baotou (*jingbao xian*) in the identification strategy since it mainly serves freight trains instead of passenger trains, which are not relevant for high-speed rail upgrade.[6,7] The five main railway lines are shown in Figure A.3 and A.4.

The validity of the instrumental variable requires two assumptions: (1) being located on the main railway lines is correlated with being affected by the high-speed rail upgrade and (2) being located on the main railway lines affects economic growth only through its impact on railway acceleration. The first assumption holds, which can be shown from the high-speed rail upgrade map in Figure A.3 and A.4. All the five main railway lines have been upgraded to high-speed rail since they connect the largest cities in China. In addition to the main

[6]One may doubt the definition of main railway lines in China. However, the definition here follows the published train schedule, where each main railway line is a section of the schedule. Therefore, the definition is objective.

[7]The railway line from Beijing to Baotou mainly serves to transported coal from Shanxi Province, the largest coal-production base in China, to other provinces.

railway lines, several regional lines have also been upgraded as shown on the map. The second assumption is somewhat stronger since the main railway lines are often located in relatively developed regions. However, even if counties located along the main railway lines are selected due to their greater economic potential, it will only bias my estimate downward as I expect the counties located on the accelerated railway lines are negatively affected in the later period of the project.

The estimation equation of a standard difference-in-difference can be expressed as:

$$\begin{aligned} Outcome_{i,t} =& \beta_0 + \beta_1 HSR_i * After_t + \gamma Year_t * Province_i \\ &+ \delta County_i + \epsilon_{i,t} \end{aligned} \tag{2.1}$$

where $Outcome_{i,t}$ is the economic outcome of county i in time t. In this chapter, I am most interested in two categories of outcome variables: (a) yearly county level GDP and GDP per capita, which represent the overall performance of a county and (b) a yearly county level investment measure, i.e., fixed asset investment, which is important because investment is a driving force of GDP growth in China (Qin et al. 2006; Yu, 1998).[8] $HSR_i * After_t$ is the difference-in-difference term, where the dummy variable HSR_i denotes whether county i was affected by high-speed rail upgrade (in 2004 and 2007) or not; and $After_t$ denotes whether it is before or after the high-speed rail upgrade for each time t. $Year_t * Province_i$ controls for year by province time trend.[9] $County_i$ controls for

[8]Fixed asset investment includes the investment in capital construction, investment in renovation and renewals of existing facilities, investment in real estate development, investment in other fixed assets by state-owned units, investment in other fixed assets by collective-owned units, private investment in housing construction as defined by the National Bureau of Statistics of China.

[9]I can also use year fixed effect instead of year by province fixed effect here if the assumption is released so that there is no heterogeneity in terms of growth trend across different provinces. The main findings do not change if year fixed effect is used.

county fixed effect. $\epsilon_{i,t}$ is the error term.

The reduced form instrumental variable estimation can be written as:

$$\begin{aligned} Outcome_{i,t} =& \beta_0 + \beta_1 Mainline_i + \sum \alpha_t Mainline_i * Year\_d_t \\ &+ \gamma Year_t * Province_i + \delta County_i + \epsilon_{i,t} \end{aligned} \tag{2.2}$$

where $Mainline_i$ denotes whether county i is located in any of the five main railway lines or not; $Mainline_i * Year\_d_t$ denotes $Mainline_i$ interacting with a series of year dummies. Other variable definitions are the same as in Equation (2.1).

It is worth noting that our sample is restricted to counties with a railroad at the beginning of our sampling period (year 1996). As county train stations also vary by size, we further exclude 98 out of 957 counties which own "large train stations" due to their historical importance in the railway system, as they might have also been considered important connections in the high-speed rail route.[10] However, the estimation results change little even if we include the counties with "large train stations."

2.3.2 Railroad data

In order to estimate the impact of high-speed rail upgrade in year 2004 and 2007 on counties being affected, I compare the economic performance of counties located on the upgraded railway lines to the counties located on conventional railway lines before and after high-speed rail upgrade, from year 2002 to 2009. Therefore, the railway status information of all the counties in China as of year

[10]Passenger train stations have been categorized into six levels according to their size and capacity, namely VIP stations and level one to level five stations. We denote "large train stations" as train stations above level three according to the standard in the 1990s.

2008 is collected from the *People's Republic of China Railroad Atlas* published in 2008. A dataset including the list of counties with access to railroad in 2008 is constructed based on the above information, along with the name of the railway line(s) on which each county is located. In addition, I identified the railroads being constructed in each year from 1996 to 2007 along with its bypass counties from the annually published China Railroad Yearbook. I excluded from the sample such counties that did not have railroad access until year 1996, since the positive economic impact of a relatively new railroad may contaminate our estimation of high-speed rail upgrade on existing railway lines.

In addition to county railroad status, the frequency of daily passenger train stops in each county during 1996-2009 is also collected for descriptive purposes. The information is manually compiled from the published passenger train schedules in each year. Each train stop is matched to its county using the *China Train Station Encyclopedia*, published in 2003. The train stops not listed in the book are matched by online resources. It turns out that only a very small proportion of train stops (around 100 out of 6,000 stops) cannot be matched to its county, as these counties are very small in size in most cases. Because those small stations are generally serviced by very few trains, this fact little affects my descriptive statistics.

Figure A.1 shows the number of operating passenger railway stations from 1996 to 2009. Around 3,000 passenger train stations were closed during the ten years of speed acceleration, especially during the high-speed rail upgrade (starting in 2004). More surprisingly, the number of counties with functioning passenger train stops is also decreasing, even with the expansion of new railway lines, as suggested in Figure A.5. Hence the number of counties that have lost train service recently has exceeded the number of counties with new access to railroads. In contrast, the accessibility of railroads in cities has slightly increased.

Figure A.6 shows the average daily train stops in each city and county during 1996-2009. It is clear that train service is much more frequent in cities than in counties. The average number of daily train stops is around 70-90 times for cities during 1996-2009, compared to merely 20-30 times for counties. Furthermore, after 2004, the average number of daily train stops indicates a decreasing trend for counties but an increasing trend for cities, in accordance with the fact that the train stations in small counties were skipped after the introduction of high-speed rail.

Figure A.7 and A.8 provide the distribution of average daily train stops in counties in 1996 and 2007, respectively. Two stylized facts can be revealed from those maps. First, the accessibility to trains is distributed unequally across counties in both years. The county with the least accessibility to railroad had only one daily train stop in 1996, while the county with the most accessibility had 345 daily train stops. However, in 2007, the county with the most accessibility to railroad service had 165 train stops, a 50 percent reduction from 1996. Second, the accessibility to trains decreased during the speed acceleration. The median of daily train services is 18 trains per day in 1996 and 14 trains per day in 2007. The two color-coded maps illustrate the decline in average accessibility to trains that accompanied the speed acceleration that occurred between 1996 and 2007.

2.3.3 County statistics data

The county statistics dataset is collected from the China Economic and Social Development Statistical Database provided by China National Knowledge Infrastructure (CNKI), which is compiled from all the publicly available statistical yearbooks and other published statistical reports.[11] All the counties and county-level cities in China have been included in the analysis except (1) counties administered by the four municipalities, namely Beijing, Shanghai, Tianjin, and Chongqing, as they are directly governed by the municipalities and are too close to the start of main railway lines, and (2) counties in Tibet, as none of them had access to railroad until 2007, which makes it unnecessary to include them in the sample based on my identification strategy. Therefore, a total of 1,878 counties are included in the sample for descriptive purposes, with information on county GDP, GDP per capita and fixed asset investment. However, only counties with train access before 1996 are included in my estimation as mentioned in Section 2.3.2. There are 957 counties for estimation purposes. The time span of the county statistics is from 2002 to 2009.

People may have some concerns about the quality of GDP data in China. However, as suggested by Au and Henderson (2006), the GDP and other economic indicators at the local level are indeed of high quality. Since our unit of analysis in this paper is at the county level, there should be little concern that the results are driven by the quality of the data.

[11] The database is available at http://tongji.cnki.net/kns55/Dig/dig.aspx with institutional access.

2.4 Findings

2.4.1 Descriptive statistics

Table B.1 shows the descriptive statistics for county level railway status and economic outcome indicators. As mentioned in the previous section, only counties with a railroad before 1996 are included. Thus, a total of 957 counties have been included in the sample, with 171 of them located on five main railway lines and 786 located on other railway lines. On average, in year 2003, around 28 trains stopped in counties located on main railway lines on a daily basis, compared to around 22 trains stopping in counties located on other railway lines. However, in 2007, both numbers dropped, from 28 to 21 and from 22 to 18, respectively. This is evidence that the reduction in train service accessibility is more severe for counties located on main railway lines than for others. In terms of economic outcomes, counties located on main railway lines on average have higher GDP, GDP per capita, and fixed asset investment. The GDP doubled from 2003 to 2007 for both groups of counties. The fixed asset investment almost tripled for both groups.

2.4.2 Difference-in-difference (OLS) estimation

Table B.2 shows the OLS regressions for the impact of high-speed rail upgrade in 2004 and 2007. Estimation results are reported for two sub-samples: 2005-2009 (which is tested for the high-speed rail upgrade in 2007) and 2002-2009 (which is tested for the high-speed rail upgrade in both 2004 and 2007). The OLS estimation using difference-in-difference specification generally suggests that the high-speed rail upgrade, especially in year 2007, hinders economic development in the affected counties. Column 1-4 of Table B.2 suggests a significant GDP and GDP per capita reduction after the high-speed rail upgrade in 2007 in the counties located on the affected railway lines, which is around 4-5 percent in magnitude.[12] However, the impact of earlier upgrade in 2004 is not significant with a negative magnitude in Column 2 and 4. The insignificant coefficient can be explained by two facts. First, the mileage of high-speed rail upgrade in 2004 is 1,960 kilometers, which is only one third of the completed upgrade in 2007 (around 6,000 kilometers.) Second, only 19 pairs of nonstop city transit trains were operating on the upgraded lines in 2004, compared to 257 CRH trains operating on the high-speed rails in 2007. Both facts illustrate that the intensity of the upgrade in 2004 is less than that in 2007. The GDP reduction is likely to be driven by a reduction of investment, as suggested in Column 5-6 of Table B.2. The decrease of fixed asset investment in the high-speed rail affected counties in 2007 is around 10-11 percent, which is doubled compared to the reduction of GDP.

2.4.3 Event study

The OLS estimation suggests that the high-speed rail upgrade in 2007 significantly hurts the economic growth in the affected counties. However, a prerequisite for the validity of difference-in-difference design is that the pre-trend of outcome variables between control and treatment groups should be similar. In

[12]The impact of high-speed rail upgrade on GDP per capita may work through its impact on population changes. However, as suggested in Table B.13, population is basically not affected in the treated counties of high-speed rail upgrade. There seems to be a one percent increase in overall population after high-speed rail upgrade in 2007 in one of the two specifications. But none of the other population measures (rural population, total number of households, and total number of rural households) is significantly affected by high-speed rail upgrade.

this subsection, I present event study graphs that plot the effects of high-speed rail upgrade in 2007 on the economic performance of affected counties. These graphs are derived from the following regression model:

$$\begin{aligned} Outcome_{i,t} =& \beta_0 + \sum_{k=-5}^{2} \alpha_k HSR_i * \mathbb{1}\{Yr_t = k\} \\ &+ \gamma Year_t * Province_i + \delta County_i + \epsilon_{i,t} \end{aligned} \tag{2.3}$$

where $\mathbb{1}\{Yr_t = k\}$ is an event time indicator equal to 1 for each year before and after the high-speed rail upgrade. Year zero is the year that the high-speed rail upgrade was implemented. For example, in year 2007, $Yr_t = 0$; while in year 2006, $Yr_t = -1$. In order to compare the effects of the event over years with the year right before the high-speed rail upgrade, year 2006 is taken as the baseline year. Therefore its coefficient ($k = -1$) is not reported in this event study. It is worth mentioning that counties that were affected by the high-speed rail upgrade in 2004 have been excluded from this analysis since the event study focuses on the upgrade in 2007.

Figure A.9 plots the event study coefficients, α_k, and 95% confidence intervals within a seven-year event window. The point estimates represent the time path of outcome variables, i.e., GDP, GDP per capita, and fixed asset investment affected by high-speed rail upgrade relative to non-affected counties conditional on county and province*year fixed effects. All three graphs support the validity of the design since none of the coefficients are significantly different from zero prior to the high-speed rail upgrade in year zero, which indicates little difference in prior growth trend between the treatment and control groups. The graphs also suggest that there seems to be a drop in GDP, GDP per capita, and fixed asset investment right after the high-speed rail upgrade, which is consistent with the previous estimation. Additionally, the negative effect is estimated to be larger as time goes by.

2.4.4 Instrumental variable estimation

In addition to the standard difference-in-difference estimation, I use an instrumental variable estimation strategy to account for the possible non-random placement of high-speed rail upgrade. More specifically, I use whether a county is located on the main railway lines to identify whether it is affected by the speed acceleration or not. Table B.3 presents the reduced form estimation following Equation (2.2), by regressing the economic indicators on the "main railway lines" dummy and its interaction with year dummies, controlling for county fixed effect and the growth trend in each province and year. It can be seen from the significance of the coefficients that the instruments, namely being located on the main railway lines over years, are good explanatory variables of GDP and GDP per capita variations at the county level. Moreover, being located on the main railway lines is a greater disadvantage for county economic growth during the high-speed rail upgrade. For example, the coefficients on $Mainline * Year07$ are negative in all of the six regressions in Table B.3, with three of them significant at the 0.01 level. The coefficients for $Mainline * Year08$ and $Mainline * Year09$ are all negative and generally larger than the coefficient for $Mainline * Year07$, which describes the trend of increasing discrepancy of GDP growth between counties located on main railway lines and other lines.

Table B.4 reports the Two-Stage Least Square (2SLS) estimation for the impact of high-speed rail on the three outcome variables. The First stage F statistic is 12.16 for the endogenous variable "$HSR07 * After$" and 71.91 for "$HSR04 * After$," which shows the strong correlation between instruments and endogenous program placement. The magnitude and significance for the 2SLS estimation is similar to the OLS estimation, except for the coefficient of "$HSR04 * After$" on the impact of GDP growth (Column 2, Row 1 of Table B.4). The same coefficient is negative without significance in the OLS regression in Table B.2. However, the coefficient is negative and significant at the 0.05 level in the 2SLS estimation, which is consistent with the significant negative coefficient on $Mainline * Year04$ in the reduced form estimation (Column 2, Row 3 of Table B.3.) In addition, the predicted average GDP growth rate for the non-affected counties and affected counties are 24.6% and 18.5% from year 2006 to 2007 according to the 2SLS estimation.[13] Therefore, the estimated impact of high-speed rail upgrade on GDP growth, which is 3.9% after translating the coefficient on log GDP into growth rate, explains around 64% of the GDP growth rate differences between the two groups.

To summarize, the findings in Table B.2-B.4 suggest that high-speed rail upgrade negatively impacts the economic growth of the counties located on the affected railway lines. More specifically, the GDP and GDP per capita of such counties decrease by 4-6 percent, which is around 336-503 million *yuan* annually, given the average county level GDP as 8.39 billion *yuan* in 2006 in the affected regions. Such negative impact explains around 64% of the GDP growth differentials between the affected and non-affected counties in the year of the high-speed rail upgrade. Furthermore, the reduction of fixed asset investment

[13]Specifically, the average predicted values of log GDP for non-affected counties are 3.65 and 3.87 in 2006 and 2007, so the growth rate is $e^{0.22} - 1$ which equals to 24.6%. Similarly, the average predicted log GDP for affected counties in these two years is 4.11 and 4.28, so the growth rate is $e^{0.17} - 1$ which equals to 18.5%. The predicted values are derived from the 2SLS estimation.

is doubled in terms of GDP reduction, which is around 10-11 percent in terms of magnitude. This can be translated as a reduction of 365-402 million *yuan* annually, given the average county-level fixed asset investment as 3.65 billion *yuan* in 2006 in the affected regions. Therefore, it can be concluded that the GDP reduction is mainly investment driven and can be explained by the drop in fixed asset investment to a large extent.

2.4.5 Robustness check using collapsed data

As mentioned in Bertrand et al. (2004), the standard error of the "treatment variable" in difference-in-difference analysis may be underestimated due to the serial correlation among the observations of the same object over years. Bertrand et al. (2004) suggest collapsing the data into "pre" and "post" periods to minimize the number of periods for each object, which helps to mitigate the serial correlation problem in difference-in-difference analysis. Following this method, I collapse the data from 2005-2009 into "pre" period (year 2005 and 2006) and "post" period (year 2007, 2008, and 2009). Similarly, I collapse the data from 2002-2009 into three periods: "pre" period I (year 2002 and 2003), "pre" period II (year 2004, 2005, and 2006), and "post" period (year 2007, 2008, and 2009).

The OLS and 2SLS estimation for the collapsed data is reported in Table B.5 and B.6. The results for OLS regressions are consistent with the estimation using disaggregated data. Similarly, the results for 2SLS estimation are consistent with the previous estimation, with the exception of one coefficient for the impact of high-speed rail in 2004. The impact of high-speed rail on GDP per capita becomes significantly positive with the collapsed data, which is contradictory to the estimation in Table B.4. One reason for this may be that the collapsed data affects the predictive power of the first stage estimation as the number of instruments decreases after collapse, which subsequently affects the second stage. It can be concluded in general that the negative impact of the high-speed rail upgrade in 2007 is more robust and consistent than the impact of the high-speed rail upgrade in 2004.

2.4.6 The impact of high-speed rail placement on cities at the prefecture level

It is shown in the previous section that less connectivity to the outside due to high-speed rail upgrade is detrimental to the small counties located on the affected railway lines. Another relevant question to ask is: have large cities benefited from better connectivity due to high-speed rail placement? It is hard to identify clearly the impact of high-speed rail on cities since they are connected to the high-speed rail "on purpose" instead of "quasi-randomly" assigned. Therefore, the identification strategy used for counties cannot be applied to the analysis of prefecture-level cities. However, in order to provide some suggestive evidence, an OLS analysis with exactly the same setting as Equation (2.1) has been conducted using all the prefecture-level cities with railroad access no later than 1996.[14]

Table B.7 (Panel A and B) shows the "correlation" of high-speed rail upgrade on prefecture-level cities in terms of both level and growth. Interestingly, high-speed rail placement does not correlate with high economic growth in the

[14] All of the GDP, GDP per capita, and fixed asset investment measures include only urban areas (districts) affiliated with the prefecture level city.

affected cities, as none of the coefficients on the double difference term are significant, though 7 out of 9 coefficients have positive signs. However, the level regressions show that GDP and fixed asset investment levels significantly increase in cities with high-speed rail upgrade, while the level change of GDP per capita does not seem to correlate with high-speed rail.

In general, the correlation analysis in cities provides some suggestive evidence that high-speed rail upgrade has only a mild impact on economic growth in the prefecture-level cities. The result, though interesting, is not very surprising for two main reasons. First, a city economy has a much larger base than a county economy. Therefore, a positive shock in transportation technology may have only a trivial impact on economic growth rate, though its impact on economic levels may not be trivial. Second, cities generally have multiple, well-developed modes of transportation networks, including not only railroad, but also highway, air, and, in the coastal areas, water. Thus, the marginal productivity increase from a technological improvement of the railway system may not play an important role. However, the marginal productivity decrease due to lost connectivity to railroad transportation is likely to be more detrimental in counties as they generally have a less developed transportation network.

2.5 Discussion

2.5.1 Heterogeneous impacts in different sectors

The main findings discussed in the above section suggest that the counties located in the high-speed rail upgrade railway lines have suffered from economic slowdown in terms of GDP, GDP per capita, and fixed asset investment compared to counties located on the non-affected railway lines. In addition, such negative impact is especially strong for the high-speed rail upgrade in 2007 compared to the early round of upgrade in 2004 due to its wider coverage with higher-lifted speed.

Since high-speed rail upgrade only affects the passenger rail services, while leaving the freight services almost unchanged, it may generate a larger negative impact on service industries (more sensitive to transportation cost of passengers) than on manufacturing industries (more sensitive to transportation cost of goods). To test this hypothesis, I estimate the impacts of high-speed rail upgrade on industrial and service sector value added in log forms following the same specifications as shown in Table B.2 and B.4.

Table B.8 reports the heterogeneous impacts of the railway upgrade on industrial sector and service sector using both OLS and IV estimation. It is suggested that the growth rate of service sector value added was reduced by 3-4 percent after the upgrade in 2007 in the affected counties. However, it seems that high-speed rail upgrade does not significantly affect the growth rate of industrial sector value added. These estimation results validate the hypothesis that the upgrade on passenger rail services affects service industries more than manufacturing industries.[15]

[15]It will be more interesting to further investigate the impacts of high-speed rail upgrade on different industries within the service sector. However, industry level GDP is not available in the statistical yearbooks. I then use the total employment in different industries in each county reported in the 2000 and 2010 China Population Census to test the impact of high-speed rail upgrade on employment changes in four industries within the service sector: hotel and restaurant; financial services; real estate and rental services. The OLS estimation results are reported in Table B.14. I do not find significant negative impacts of high-speed rail on employment changes in these four industries from year 2000 to 2010, even though three out of the four coefficients are negative. The estimation would be more precise if yearly industry level data is available.

2.5.2 Impacts by distance to the nearest high-speed train station

Counties are not equally distant to high-speed train stations in the urban areas. Some are close to the urban districts, while others are a few hundred kilometers away. It is thus interesting to examine whether the negative impact of high-speed rail upgrade varies by the geographical proximity to high-speed train stations. On the one hand, it is possible that the counties close to the urban core were negatively affected the most since more investment was diverted from such counties to the well-connected urban areas due to proximity. On the other hand, it is also be possible that the counties distant from the urban areas were hurt the most since positive agglomeration spillovers from the cities to the nearby counties may offset some of the negative impact in these counties close to high-speed train stations.

In order to test the possible heterogeneous impacts by distance to high-speed train stations, I compute the distance (unit:100 km) from the centroid of each county to the centroid of its nearest city with high-speed train stations, and interact the distance and squared distance with the difference-in-difference coefficient to estimate the possible heterogeneous impacts. Table B.9 shows the results. "$HSR07 * After * Distance$" and "$HSR07 * After * DistanceSquared$" are the two triple difference terms. In addition to these terms, I also control for each pairwise interaction and the main effects. The results suggest that the impact of high-speed rail upgrade does not vary by the proximity to high-speed train stations in the urban areas. None of the triple difference terms are significant at the 0.1 level. The coefficient on "$HSR07 * After * Distance$" is marginally negatively significant for fixed asset investment, indicating some weak evidence that the negative impact of high-speed rail on investment increased by distance.

2.5.3 Channels

There are two possible channels which may lead to economic slowdown in the high-speed rail affected counties. First, since high-speed trains squeeze out some of the conventional train services in the affected counties, the train accessibility decreases in those counties, which implies a transportation cost (or trade cost) increase in such places. As a consequence, the affected counties will become less integrated and economic activities will decline in response to the increase in trade barriers. Second, high-speed rail upgrade connects large cities more tightly by reduced commuting time, which intensifies the agglomeration forces between large cities. In that case, it is likely that the economic activities will divert from small counties to large cities in order to enjoy the agglomeration spillovers generated from a more tightly connected transportation network. This possible channel resonates with the findings in Faber (2013) that the national trunk highway system in China leads to reduced industrial output in connected counties, given no reduction in transportation cost in those counties.

As mentioned in Section 2.2.2, a county being "affected" by high-speed rail, i.e., located on the high-speed rail upgrade lines, does not necessarily have a train service reduction after the rail upgrade since a county's overall train services are also determined by the services provided by other railway lines that pass through the county. Therefore, in order to test the validity of the first channel, I further examine the heterogeneous impact of high-speed rail upgrade in the affected counties that suffered from train service reduction in the year 2007 (group A) and the other affected counties that did not experience train service reduction during the same period (group B). If the increased trade cost due to train service reduction is a channel for the reduced economic activities in the affected counties, the impact of high-speed rail in group A should be more negative than that in group B since group A suffered from a larger increase in trade cost.

Table B.10 presents the comparison of high-speed rail's impact between the above-mentioned two groups. The variable "*train service not reduced*" equals to one if the observations belong to group B, otherwise zero. While the interaction term between the double difference coefficient ($HSR * After$) and the trade cost dummy variable (*train service not reduced*) indicates the difference of high-speed rail's impact between the two groups. It is shown that two out of the three interaction terms are positive, which works in favor of our hypothesis that the impact of high-speed rail upgrade on group B is less negative than that on group A. However, the difference in terms of impact is not statistically significant which provides weak support to the first channel.

In order to test the second channel, i.e., whether the improved agglomeration benefits between large cities after high-speed rail upgrade diverted economic activities from small counties to large cities, I collected the highway status of all the counties in the sample before and after the high-speed rail upgrade in 2007. The hypothesis is that, if a county has already been connected to the highway network prior to high-speed rail upgrade in 2007, some of its economic activities have already been diverted to the large cities on the highway network (Faber 2013). Therefore, if the second channel works, the high-speed rail upgrade's negative impact on diverting economic activities away should be smaller comparing to its impact in counties that had no highway network prior to high-speed rail upgrade.

Table B.11 displays the comparison of high-speed rail's impact between counties with and without highway access prior to high-speed rail upgrade in 2007. The dummy variable "$connected\ to\ highway\ before$ 2007" equals to one if the county was connected to the highway network before year 2007, zero otherwise. Similar to the test of the first channel, an interaction term between double difference coefficient and highway status ($HSR07 * After * Connected\ to\ Highway\ before$ 2007) is included in the regression to test the differential impact in counties with different highway access. It is shown in Table B.11 that all the three interaction terms have a positive coefficient, indicating that counties with highway access prior to 2007 suffered less from high-speed rail upgrade than counties without highway access. Especially, the differential impact between the two groups is most significant for GDP in terms of both magnitude and significance. High-speed rail upgrade reduced GDP by only 3 percent in counties with highway access, while the impact in counties without highway access was three times larger (9 percent). Therefore, the second channel is likely to play a role in explaining counties' reduced economic activities due to high-speed rail upgrade.

2.5.4 Magnitude of the impact

The main results suggest that high-speed rail upgrade negatively impacts the GDP growth rate of the counties located on the affected railway lines by 4-5 percent. Answers to the following two questions may help us better understand the magnitude of such impact. First, how large is the impact compared to the average economic growth rate in those areas? The annual GDP grew from 4.38 billion *yuan* in 2002 to 13.65 billion *yuan* in 2009 in the affected counties, with an overall growth rate of 311.6%, translating into an annual growth rate of 17.6%.

This implies that the magnitude of 4-5 percent GDP reduction is economic significant for the economic growth of the counties affected by the upgrade.

Given that high-speed rail upgrade leads to significant economic slowdown in the affected counties, the next question to ask is whether the GDP reduction in counties outweighs the economic gains in cities, which makes high-speed rail upgrade an unattractive investment in terms of its economic returns. Given that the average GDP for the 183 affected counties is 8.39 billion *yuan* in 2006, the total loss of GDP in the 183 counties is 76.77 billion *yuan* in 2007.[16] A total of 80 cities have been connected with high-speed rail in 2007, thus the net economic return of the investment in its first year would be positive as long as the average economic benefit in cities exceeds $76.77/80 = 0.96$ billion *yuan*. Since there is no credible estimate on the causal impact of high-speed rail upgrade on GDP growth in cities, I use Panel B of Table B.7 as a possible benchmark, where the estimated benefit of the upgrade is 16.76-21.67 billion *yuan* of GDP increase in the affected cities. Based on that, the benefit that high-speed rail upgrade brought to cities seems to be more than enough to compensate for the losses in counties.

2.6 Conclusion

Infrastructure is supposed to promote economic growth. However, infrastructure investment with a preference in the urban sector may generate negative externalities to the less developed rural sector in developing countries. The quasi-experiment of high-speed rail upgrade in China in the year 2004 and 2007

[16] I use 5% here as the maximum negative impact of high-speed rail upgrade.

provides an ideal case to study the distributional consequences of infrastructure investment.

Applying conventional difference-in-difference (OLS) and instrumental variable strategies, I come out with the following main conclusions. First, by comparing GDP and GDP per capita of counties located on the affected railway lines to counties located on other railway lines, evidence suggests that there is a 4-6 percent significant reduction in GDP and GDP per capita after the 2007 high-speed rail upgrade in the counties located on the affected railway lines. After accounting for nonrandom high-speed rail placement, the 2SLS estimation shows consistent estimation in terms of significance and magnitude. Such impact could explain around 64% of the predicted GDP growth differentials between the affected and non-affected counties right after high-speed rail upgrade. Second, the GDP reduction in the high-speed rail bypassed counties, which is around 336-503 million *yuan*, given the average county level GDP as 8.39 billion *yuan* in 2006, can be largely explained by the concurrent drop in fixed asset investment. Third, since high-speed rail upgrade affects the transportation of passengers and not transportation of goods, its negative impact is more pronounced in the service sector than in the manufacturing sector. Lastly, the diverted economic activities from small counties to large cities due to increased agglomeration forces is likely to be a channel accounting for the negative impact of high-speed rail. Together, these results imply that the distributional implications of these types of investments can be dramatic.

In addition to the negative consequences brought to counties, it seems that the introduction of high-speed rail upgrade in prefecture-level cities is only associated with mild positive impact in terms of GDP and investment. However, it is likely that some of the benefits could not be captured given the limited availability of data, such as its potential positive impact on subjective welfare of urban residents. More importantly, it may generate some long-run impact in reshaping the economic geography of the country, which may not be captured for now. Furthermore, future research is needed to establish a cleaner causal relationship between high-speed rail and economic growth at the prefecture-city level.

CHAPTER 3

THE ROAD TO SPECIALIZATION IN AGRICULTURAL PRODUCTION: EVIDENCE FROM RURAL CHINA

3.1 Introduction

In developing countries, many rural poor live in isolated areas. Because they reside far from markets, the poor are more likely to rely on self-sufficient, subsistence farming to survive. Spatial poverty traps are a silent feature of the rural landscape (Jalan and Ravallion 2002). Scholars have argued that rural roads are a key instrument in overcoming spatial poverty traps in developing countries (Caldern 2009; Escobal and Ponce 2002; Fan and Hazell 2001; Jacoby and Minten 2008). However, rural roads may be costly to build, and therefore rigorous impact assessments of the effects of rural roads in lagging areas are necessary before policy interventions.

A limited number of studies have evaluated the returns to investing in roads in developing countries, but many of them are conducted at the aggregate level (Fan and Hazell 2001; Fan and Zhang 2004; Zhang and Fan, 2004). Those studies have been criticized for failing to uncover the mechanisms by which road connections shape household production and consumption behavior (Jacoby 2000). Studies at the household level, on the other hand, often rely on cross-sectional data due to difficulties in obtaining long-term time series data in poor areas. However, cross-sectional data cannot address the problem of endogenous road placement, that is, roads are more likely to be built in high-potential areas. To overcome the problem of endogeneity, Jacoby (2000) develops an innovative approach to evaluate the impact of road access on agricultural land value, computed based on the discounted stream of maximal profits from cultivation. Yet the approach is inadequate for evaluating impact on the welfare of landless laborers, who are common in developing countries. In the context of China, the method is also inapplicable because farmers do not own the land but only hold the right to cultivate it. In the absence of agricultural land markets, uncovering true farmland value proves difficult.

In this chapter, we use a primary panel household dataset collected in a remote and poor area of China to investigate the impact of road connections on rural welfare by focusing on agricultural specialization and input use. Road connections can potentially reshape the production choice set of isolated farmers and affect agricultural production, the major livelihood of the poor, in at least two ways.

First, with lower transportation costs, farmers may shift their agricultural production from autarkic, subsistence farming to more market-oriented, specialized activities (Limao and Venables 2001; Renkow, Hallstrom, and Karanja 2004). Yang and Ng (1993) develop a theoretical model showing that producers will choose to specialize in one activity according to their comparative advantage and simply purchase other goods and services from the market, provided that transaction costs are sufficiently small. In contrast, when transaction costs are too high, it makes more economic sense for producers to remain autarkic. Using a simulation approach, Omamo (1998) finds that as distance to the market shortens, small-scale farmers tend to shift away from diversified cropping patterns in favor of cultivating only one crop. However, the empirical findings are mixed. For example, Stifel, Minten, and Dorosh (2003) show that in Madagascar, the concentration level of agricultural production in the least remote areas is around 1.5 times that of the most remote areas, suggesting that improved road access facilitates specialization in agricultural production. Gibson and Rozelle (2003) provide a counterexample: they find that in Papua New Guinea, each extra hour it takes to reach the nearest road induces a 2.6 percent reduction in the number of activities, in contrast to the theoretical prediction. However, the variable "number of activities" does not necessarily reflect the intensity of each activity, such as the time spent, income earned, or area cropped. Therefore, the result in Gibson and Rozelle (2003) may not be in direct conflict with the estimation by Stifel, Minten, and Dorosh (2003) from the dimension of specialization.

Second, as improved road access reduces transportation costs, the prices of modern inputs such as fertilizer are more likely to drop (Khandker, Bakht, and Koolwal 2006). Consequently, farmers may apply more modern inputs to improve agricultural productivity. In addition, farmers may hire more labor to take care of specialized agricultural production as road access improves. Gollin and Rogerson (2010) develop a theoretical model and calibrate it with Ugandan data, showing that as transportation cost declines, farmers will use more intermediate inputs, which in turn contribute to agricultural output growth. The empirical findings on the impact of rural roads on modern input use, however, are inconclusive. Benziger (1996) finds that better road access leads to increasing fertilizer use in villages in Hebei, China. Stifel, Minten, and Dorosh (2003) show that farmers in more isolated regions of Madagascar use less fertilizer than those in places with better road access. However, Dorosh et al. (2010) paint a more complicated story: input use depends on not only distance to roads but also the density of road networks. In East Africa, for example, reducing travel time significantly increases adoption of high-input/high-yield technology, whereas roads have an insignificant impact in West Africa, where road network density is relatively higher at the beginning of the sampling period.

One challenge to an empirical evaluation of the impact of road access on agricultural production is data limitation. Most empirical studies rely on cross-sectional data, making it hard to control for unobserved factors, such as the placement effect mentioned earlier. In this chapter, we use a primary household panel dataset collected in 17 natural villages over four waves in Guizhou Province, China, to investigate how road access shapes farmers' cropping decisions and their livelihoods.

Our dataset possesses two advantages when studying the impact of access to road networks in isolated villages. First, given that it relies on non-recall panel data, our study provides relatively accurate and credible information with respect to household agricultural production. Second, the four waves of data allow us to conduct a difference-in-differences analysis, which helps mitigate estimation biases as a result of omitting variables and reverse causality commonly seen in regressions based on cross-sectional datasets. To the best of our knowledge, this is the first paper to empirically document the impact of road access on agricultural specialization and input use in China.

We find that access to roads fosters household agricultural specialization. The impact is economically significant and is about one-fifth of the standard deviation of the Herfindahl-Hirschman specialization index (HHI). In addition, better road connectivity induces farmers to apply more fertilizer and spend more money hiring labor. Thanks to those two channels, road access is shown to boost farmers' agricultural income, which in turn contributes to poverty reduction. However, the introduction of road access does not seem to improve farmers' nonagricultural income in this remote area.

The findings may have some policy implications for China. In the past several decades, the Chinese government has made significant investments in building a nationwide highway system. As the highway density increases, the marginal returns to highway investment are likely to decrease. Fan and Chan-Kang (2005) argue that it may make more economic sense to gear investment toward rural roads. But rural roads carry less traffic, are harder to maintain, and are more costly to build in remote areas. Therefore, it is important to gather more empirical evidence as to how rural roads affect agricultural patterns and rural livelihoods in lagging regions.

One should be cautious in explaining the findings. Our sample focuses only on the mountainous rural areas in southwestern China, where smallholder farming is the dominant mode of agricultural production. As China is a large and spatially diverse country, the findings drawn from this sample may not apply to China as a whole. Our study is more relevant for understanding as to how road connections might affect farming practices and rural livelihoods in isolated and impoverished regions.

3.2 Description of Data

As Figure A.10 shows, Guizhou is located in southwestern China. Guizhou is one of China's poorest provinces and has the shortest road length due in part to its mountainous terrain.[1] Figure A.11 depicts the road system in Guizhou as of

[1]According to China Statistical Yearbook (2005), GDP per capita in Guizhou is 4,317 yuan in 2004, lowest among all the mainland provinces in China. Highway and level I road length in Guizhou is also the lowest among all the mainland provinces as of year 2004.

2004. Highway networks are sparse in Guizhou, with only four reaching from the provincial capital (Guiyang Shi) to major cities in the province. Although national and provincial roads are numerous, the density is much lower than the national average. In remote mountainous villages, some households still practice subsistence farming, whereas households in relatively flat areas sell most of their agricultural products to the market. The large variation in road access in our sample thus provides us with a valuable opportunity to study the impact of road access on agricultural production in isolated regions.

The survey site, Puding County, comprises 11 townships and 317 administrative villages, and as of the end of 2008 had a total population of 448,000 people.[2] A highway and a national road bypass the county border, and one provincial road cuts through the county. In 2008 the average household income in Puding County was around 5,800 yuan, slightly above the provincial median but below the provincial mean. [3] As Figure A.12 depicts, in terms of per capita rural income, Puding is in the middle tercile, suggesting Puding is a rather representative county in Guizhou Province.

Three administrative villages representing different levels of economic development of Puding were chosen for the survey. The three administrative villages (henceforth referred to as Administrative Village I, II, and III) contain 17 natural villages. A census-type survey of households in all the natural villages was first administered in early 2005 and included 805 households. A second sur-

[2]An administrative village is a bureaucratic entity comprising several natural villages (hamlets). A typical natural village includes 30 to 50 households. It is too small to form an administrative unit. As a result, some nearby natural villages are artificially put together to create an administrative village. However, in the mountainous area, it sometimes takes one a few hours to walk from one natural village to another within the same administrative village.

[3]In our sample, the average household income in 2006 is 7,619 yuan, which is above the mean household income according to official statistics. It is likely that our sampled township is close to the county seat which is richer than the county as a whole.

vey wave, covering 833 households, was conducted in early 2007. A third wave was undertaken in early 2010 and surveyed 873 households. And the fourth wave was carried out in early 2012 covering 943 households. [4] The surveys collected detailed information on household characteristics, demographics, income, agricultural production, and consumption.

The natural villages vary widely in their degree of road access. We define a road as being accessible if tractors can drive through during the rainy season. [5] Using information collected from the records of village offices, Table B.15 summarizes road access in the 17 natural villages in 2004, 2006, 2009, and 2011. Administrative Village III, which is right next to the county seat, has the best road access of the three administrative villages. All of Administrative Village III's natural villages already had road access prior to the first wave of the survey. Four natural villages in Administrative Village I constructed roads during our survey periods. However, until our most recent survey, some natural villages, such as Natural Village 1 and Natural Village 3, had yet to gain road access. In Administrative Village II, one natural village built a new road during 2004 and 2006, whereas two other natural villages still lacked road access at the time of our most recent survey.

As we are interested in the impact of road access on agricultural specialization, we constructed a Herfindahl-Hirschman index as a measure of specialization at the household level. The HHI is defined as the sum of the squares of agricultural income shares derived from different production activities.[6] The

[4] The total number of surveyed households varies across different years due to cases of families split, migration, and a few attritions. There are 782, 815, 834, and 935 valid observation households in the four waves, respectively.

[5] Market activities also exist during rainy season. Therefore it is necessary to emphasis the constraint of "rainy season" in the definition of road.

[6] For example, if the household produces maize and fruit, with an income (including in-kind) of 2,000 yuan and 3,000 yuan, then the specialization index is calculated as $(2000/5000)^2$ +

specialization index ranges between 0 and 1. The greater the value, the higher the degree of specialization.

Table B.16 reports income sources from several major agricultural activities. For each agricultural activity, total income is the sum of cash income and in-kind income (imputed using market price of the specific year). Maize is the predominant crop, generating the largest share of agricultural income, ranging from 37 to 46 percent in the four survey years. As the second most important crop, rapeseed provides 16 to 21 percent of household agricultural income. Livestock ranks third in terms of agricultural income generation. In our sample, approximately 18 to 30 percent of households were engaged in livestock production compared to approximately 90 percent participation rates in maize and rapeseed production.

It is worth noting that the categories of agricultural income decomposition are slightly different across the three waves of surveys due to changes in questionnaire design. For example, the 2004 and 2009 waves contain nine subcategories of agricultural income, whereas the 2006 and 2011 waves have 10 subcategories. Additionally, no data are available for vegetable income in 2009. To address these problems, we construct two alternative specialization measures as robustness checks based on different classifications of income categories. For the first alternative measure, we impute the vegetable income in 2009 based on the actual vegetable seed cost available in 2009 and then estimate the past relationship between vegetable seed cost and income observed in the first two survey waves (variable denoted as HHI [2]). In so doing, we obtain comparable household vegetable income for all the four waves. For the second alternative measure, we reclassify the non-overlapping subcategories and the rest of the in-

$(3000/5000)^2 = 0.52$.

come as "other." After the adjustment, there are nine comparable subcategories of agricultural income across the four waves (variable denoted as HHI [3]).

Table B.17 presents the summary statistics for the key variables used in the analysis.[7] Average household income more than doubled from 6,246 yuan in 2004 to 16,538 yuan in 2011. Income generated from nonagricultural activities played a key role in overall income growth. Nonfarm income grew from 2,267 yuan in 2004 to 10,840 yuan in 2011. By comparison, average household agricultural income grew at a slower pace, from 3,978 yuan to 5,698 yuan, during the seven-year period. The relatively lackluster performance in the agricultural sector is not surprising given limited arable land in this area. After all, Guizhou ranks among the lowest in per capita arable land in the Chinese provinces. On average each person in our survey village cultivated only 0.81 mu of land in 2011, about half of the national average of 1.4 mu per capita.[8]

Lastly, the mean level of household agricultural specialization index is 0.46, 0.41, 0.49, and 0.47 in 2004, 2006, 2009, and 2011, respectively. The drop in the specialization index in 2006 is perhaps due to that year's severe drought. In 2006, the share of corn income dropped to 39 percent, lower than that of 2004 (46 percent) and 2009 (42 percent). The drought may thus result in the blip in the trend of the HHI.

The summary statistics reveal stark differences between households with and without road access. As Table B.18 shows, the mean household income in villages with road access is almost double that of villages without road access. Both the agricultural and nonagricultural incomes per capita in households with road access are higher than the incomes of their counterparts. In terms of agri-

[7] All the income and price measures have been deflated to year 2004.
[8] 1 mu = 0.066667 hectare.

cultural production, the villages with roads were more specialized than those without access to roads. In general, households with road access tend to be of a smaller size, have larger areas of cultivated land, and have higher levels of education.

3.3 Empirical Model

Our empirical question is the following: does road access have any impact on the extent of specialization and input use in agricultural production? In this paper, we adopt a difference-in-difference method to answer that question. The specification is as follows:

$$Y_{i,t} = \alpha_0 + \beta_1 Road_i * Beforeafter_{i,t} + Z_{i,t} + \phi_{village} + \psi_{year} + \epsilon_{i,t} \quad (3.1)$$

where $Y_{i,t}$ is a dependent variable for household i in time t; $Road_i$ denotes whether the village to which household i belongs has a road by the end of our last survey wave (year 2011); $Beforeafter_{i,t}$ denotes whether the village household i belongs to has road access in year t; $Z_{i,t}$ represents a series of control variables, including cultivated land area, number of primary-age laborers (being from 16 to 60 years old) in a family, household size, the highest year of schooling within the household, and whether there is a village leader in the household; $\phi_{village}$ stands for natural village fixed effects; ψ_{year} controls for year fixed effects; $\epsilon_{i,t}$ is the error term.

Our coefficient of interest is β_1, the double difference term, which represents the impact of road access on the outcome variables. The main dependent variables in our estimation are (i) the household agricultural specialization index (HHI [1], HHI [2], and HHI [3]); (ii) fertilizer use measured by the natural log of the monetary value of fertilizer use per mu of land; (iii) expenditures on hired labor in logarithmic form; and (iv) the natural log of agricultural income, nonagricultural income, and total income per capita in the household. If road access promotes agricultural specialization, we expect β_1 to be positive and significant. Similarly, β_1 is expected to be positive and significant as well if the outcome variable is either fertilizer use, the cost of hired labor, or household income.

Because we have a panel dataset, we can largely remedy the common problems plaguing cross-sectional analyses. For instance, we can include household characteristics, natural village fixed effects, and year fixed effects to mitigate omitted variable bias. Since the cropping and input use decisions depend upon the existing road conditions, instead of specialization and input use on road placement, reverse causality is unlikely. It is also hard to imagine farmers would change their cropping patterns in anticipation of a new road in the next several years. Perhaps the biggest challenge is road placement. As suggested by recent impact evaluation literature (Duflo and Pande 2007), the nonrandom program placement may bring about endogeneity problems in economic estimations. A typical solution is to carry out a two-stage least-squares analysis by instrumenting the policy with a set of exogenous variables. However, the road variable varies only at the natural village level and there are only 17 natural villages in the dataset, making it impossible to implement the first-stage regression with such a small number of observations. Since our objective is to examine how households respond to road connections in their production decisions, the potential endogeneity problem of road placement, if any, is minimal.

Since the road status change is at the natural village level, we need to cluster the standard errors by 17 natural villages. However, when a sample comprises of a small number of clusters (fewer than 30), the conventional double-difference estimation on the standard errors may become less precise (Cameron, Gelbach, and Miller 2008). More specifically, the clustered standard error tends to over reject the null hypothesis. Therefore, we adopt the wild cluster bootstrap-t procedure, which Cameron, Gelbach, and Miller (2008) have demonstrated as having good size properties with small number of clusters, and report the bootstrapped p-values for the key variable of interest in the regression tables. [9]

3.4 Empirical Results

Table B.19 reports the main regression results on specialization, fertilizer use, and cost of hired labor. Natural village fixed effects and year fixed effects are included in all the regressions to control for village-specific factors, such as village growth potential and common temporal trends such as investment policy. The first column under each heading lists the most parsimonious specification, and household characteristics are added in the second column.

For specialization, we use three indexes: HHI [1], HHI [2], and HHI [3]. Regardless of the two different specifications, road access is shown to have posi-

[9] Appendix B of Cameron, Gelbach and Miller (2008) introduces the details of the wild bootstrap-t procedure. Basically the bootstrap procedure resamples residual using Rademacher weights (equal probabilities of 1 and -1) to obtain a new sampling of residuals from a restricted regression with a null hypothesis ($\beta_1 = 0$ in our model). The Wald statistic of the OLS estimation with clustered standard error is calculated for each pseudo-sample. The bootstrapped p-value is inferred from the location of the original Wald statistic in the distribution of bootstrapped Wald statistics in 999 replications.

tive and significant impact on agricultural specialization. The results are robust to three slightly different specialization indexes. On average, road access improved the specialization index by 3 percentage points, which is approximately one-fifth of one standard deviation of the HHI.

As Table B.19 shows, better road connections also induce farmers to apply more fertilizer. After improvements in road connections, fertilizer use (yuan per mu) rose by 33.6 percent (after translating the log form coefficient 0.29 into a real growth rate). Similarly, road access boosted household expenditure on hired labor (yuan per mu) by 75.1 percent (after translating the log form coefficient 0.56 into a real growth rate).

As farmers specialize in their agricultural production, apply greater amounts of modern inputs, and hire more skilled professional workers, we expect their agricultural income to increase as well. Table B.20 summarizes the regressions on household income, including agricultural income, nonagricultural income, and total income per capita. There is some evidence that road access enhances agricultural income by 27.1 percent (after translating the log form coefficient 0.24 into a real growth rate), which is marginally significant with a bootstrapped p-value of 0.15. However, roads do not appear to play a major role in shaping nonagricultural household income. In this area, most young people migrate outside of the province to work in the nonfarm sector. Road conditions are not a binding factor to their migration decision. Overall, the impact of roads on total income is positive but not significant.

The identification assumption for difference-in-differences requires parallel growth trend between the control and treatment villages prior to road placement. However, we cannot directly test this assumption since 1) our data contains only four data points for each observation, and 2) roads were placed in various years instead of at the same time. In order to shed some lights on the robustness of the main results without the most direct evidence on common trend, we conduct a placebo test by moving the road placement to one survey period earlier. For example, if the village was not given access to road until the second wave of survey, we take this village as being treated with road placement since the first wave, instead of the second wave, in our placebo treatment.

Table B.21 shows the results for placebo test. The sample is restricted to observations without access to rural roads, otherwise the estimation results will be contaminated by the improved outcomes in the affected households after the introduction of roads. The impacts of a placebo treatment of rural road, which is one period before the real treatment, on five outcome variables are reported in this table, including three different measures of specialization indices and two measures of input use. They are the outcomes which have proved to be significantly affected by road status in the main estimation results. However, neither coefficient is significantly different from zero in the placebo test, which provides robust evidences on the validity of the difference-in-differences identification.

Considering that most of the poor still depend on agricultural production as their major livelihood and having shown that better road connections help farmers improve agricultural income, naturally we expect that road development facilities poverty reduction. To test that hypothesis, we regress three common poverty measures (P0, P1, and P2) at the natural village level on the following variables:[10] a dummy variable indicating whether a natural village has

[10]P0 measures poverty incidence (the proportion of people living under the poverty line). P1 (the so-called poverty gap index) measures the gap between the actual income and the poverty line. P2 averages the squared poverty gaps relative to the poverty line, which implicitly attaches greater weight to the poorer segment of the population in the measurement. See Foster, Greer, and Thorbecke (1984) for details.

road access in 2011 or not, its interaction term with a dummy variable for the years with road connection, acreage of land, primary-age population, presence of a village cadre, and year fixed effects. The poverty measures hinge crucially on the definition of the poverty line. Under a low poverty line, fewer people will be counted as poor, while using a high poverty line entails a higher poverty incidence. To check the robustness of the results to the choice of poverty line, we calculate two sets of poverty measures based on the official Chinese poverty line and the international $1-per-day poverty line. The official poverty line is 668 yuan in 2004 prices, equivalent to only $0.66 measured in 1985 purchasing power parity (see Xing et al. 2009). Using the international poverty line of $1.08 per day per capita, the poverty line in China in 2004 would be 892 yuan.

Table B.22 presents the regression results. Panels A, B, and C show the key variable of interest, the difference-in-differences interaction term for the three dependent variables, P0, P1, and P2, respectively. Under each panel are two sets of regression results, one for the low poverty line and one for the high poverty line. Under the heading of low poverty line or high poverty line, we further present two different specifications: no village fixed effects, and with administrative village fixed effects. Since our panel dataset is at the natural village level, in principle we should include natural village fixed effects to control for unobserved natural village specific factors. However, since the poverty measure is much less variable than the income measure and the number of observations is rather limited, including natural village fixed effects likely would wipe out the variations of the dependent variables. Therefore we do not include natural village fixed effects in the regression tables.[11]

[11] The AIC (Akaike information criterion) shows exactly that point. In effect, the most parsimonious regressions without any fixed effects have the lowest AIC, providing the best fit to the underlying data generation process. In contrast, the models with the natural village fixed effects perform the worst.

Between the two preferable specifications, the parsimonious regressions and those including only administrative village fixed effects, the coefficient for the difference-in-differences interaction term is generally significantly negative regardless of the choice of poverty measure, suggesting road development contributes to poverty reduction. The channel of impact is likely through increased agricultural production and income that the poor primarily rely on.

3.5 Mechanisms

The baseline result in the above section shows that rural road access significantly improves the specialization level of agricultural production, as measured by income-based Herfindahl-Hirschman index. In principle, road access affects the income-based specialization index through two possible mechanisms: (1) reallocating cropping areas; and (2) applying more modern inputs (such as fertilizer) thanks to lower transportation cost, which in turn boost yield. Since farmers in these remote villages are largely price takers, rising yield naturally results in a greater share of crop income.

In order to probe into the two potential mechanisms, we first calculate the area-based specialization index at the household level in each survey year.[12] Poultry, livestock and fishing activities are not included since they are not area based. In the next step, we run similar regressions based on Equation (3.1) to study the impact of road access on area-based specialization index. As shown in the first two columns of Table B.23, the coefficient on area-based HHIs is positive and marginally significant, providing some weak evidence that road

[12]The households in the second wave (year 2007) are excluded since there is no crop specific area data.

access promotes the specialization of crop production.

To investigate the second mechanism, we compute the yield of maize and rice, the two most popular crops in the area, respectively.[13] Again, due to the lack of crop specific area data, we are not able to compute the crop yield for the households in the 2007 survey. Therefore, only observations in the first wave and the last two waves have been included in the analysis. Column 3 6 of Table B.24 shows the results for the impact of road access on the yield of maize and rice. The yield of maize has been significantly increased by around 45 jin per mu after road is introduced to the natural village. However, the impact of road on rice yield is positive but insignificant. The results are consistent with our observations in the field. Each household can have multiple plots. In general, rice is mainly produced in relatively plain areas, normally with good road access in history. So we don't expect to find new access to road has significant impact on rice yield. However, maize is mainly produced in hilly areas. So improvement in road access lowers the transportation cost of fertilizer, inducing farmers to apply more fertilizer which boosts yield and income.

In a word, as road access improves, farmers are likely to focus on a few numbers of crops and apply more modern inputs, resulting in higher yield for the chosen crops. Consequently, farmers' agricultural income becomes increasingly concentrated in a few numbers of crops.

[13]The unit is defined as jin per mu, where 1 jin = 0.5 kilogram.

3.6 Conclusion

In this chapter, through the use of primary census-type household surveys in remote villages in China, we examine the impact of road access on agricultural production, particularly on specialization, and intermediate input use. We find that better access to roads facilitates farmers to specialize in agricultural production, induces them to use more fertilizer, and prompts the hiring of more laborers. Putting those factors together, road access is shown to promote agricultural income and contribute to poverty reduction. However, its impact on nonagricultural income is rather minimal. There are two potential reasons for the insignificant impact on the nonagricultural sector. First, the area is rather remote. Even with improved road access, local rural nonfarm activities are still rather limited compared with the coastal regions. Second, the rise in real wages as a result of the arrival of the Lewis turning point (the exhaustion of surplus labor) since the mid-2000s has attracted a larger number of rural workers to cities (Zhang, Yang, and Wang 2011). Under such circumstances, farmers increasingly rely on remittance as the major nonfarm income. In this remote area, farmers' migration decisions may have little to do with local infrastructure conditions.

In this chapter we find that road access helps facilitate the market integration of the agricultural economy, therefore enlarging the production scale of products with comparative advantage. For example, in Natural Village 4 of Administrative Village II, the natural endowment is suitable for growing peaches. Before improvements in road connections, peaches were often damaged after being carried by shoulder for a long walk to the nearest market. After road construction, farmers can sell their peaches at a collection point right in their natural village. As a result, peach production has boomed in this area. Given the limited data that we have, we try our best in this paper to rule out the other possibilities which may contaminate the difference-in-differences estimation by applying a placebo test and using conservative estimation of standard errors. Across all the settings, the evidences consistently suggest that in areas with road investment, we tend to observe an increasing trend of agricultural specialization. However, we admit that we may not fully address the endogeneity concern of road placement given the available datasets.

Also, we shall caution that the findings on road investment's positive impact on agricultural production do not necessarily mean that roads should be built connecting all the remaining natural villages, as the marginal cost of building roads to the more remote communities may far outweigh the benefit. Thus, a cost benefit analysis is needed when considering such rural road projects.

CHAPTER 4

INSECURE HOME OWNERSHIP AND RISING SAVINGS RATES IN CHINA: EVIDENCE FROM REPORTED FORCED EVICTIONS

4.1 Introduction

Savings rates in China have increased a lot in the recent decade. Regardless of government and corporate savings, the Chinese household savings itself as a share of disposal income nearly rose from 16 percent in 1990 to 30 percent in 2007 which caught special attention (Wei and Zhang, 2011). There are two prevailing explanations with well-founded empirical evidences for the rising household savings rates in China. First, precautionary savings motive due to less provision of education, health and housing services, in combination with a rise in income uncertainty may drive up household savings in China (Blanchard and Giavazzi 2006; Chamon and Prasad 2010). Second, demographic changes, such as sex ratio rises in China may lead to higher household savings rates since Chinese parents with a son raise their savings in a competitive manner in order to improve their sons relative attractiveness for marriage (Wei and Zhang, 2011). There are several other explanations as well, such as low level of financial development and cultural norms.

In this chapter, I study whether the increasing insecurity of home ownership being reported in the media induces urban households to save more in China, which is in favor of the precautionary savings motive. This is a less studied question with increasing importance. On the one hand, home ownership in China is associated with reduced household savings. Chamon and Prasad (2010) find that owning a home is associated with sharply lower savings rate

(4-7 percentage points) among young households. It is thus interesting to investigate whether owning a home with uncertainty of being demolished will induce households to save more. On the other hand, insecurity of home ownership is a growing social problem in China in recent years along with the rapid growth of infrastructure investment and urbanization process. However, there is no paper studying the potential consequences of such rising social problem in China as far as is concerned.

I collected the incidence of forced evictions being reported in the largest Chinese news search engine, Baidu News Archive, in each city and year during 2004 to 2011, combined with the average urban household savings rate constructed from the average disposable income and expenditure data from CEIC database in prefecture level cities. In addition, I also collected information on other economic and demographic indicators at the city level from China Economic and Social Development Statistical Database.

Using dynamic panel data models, I find that worse insecurity of home ownership, as indicated by more frequent forced evictions, leads to higher household savings rate at the prefecture city level. To be specific, if the incidence of forced eviction is increased by 161%, which is the average annual growth rate suggested from the data, household savings rate in that city is expected to increase by 0.64 to 0.84 percent. I find that there are two possible mechanisms which may lead to this result. First, it is likely that the impact of forced evictions on household savings rate works directly through a reduction in home sales. If forced evictions discourage home sales, it may increase the average household savings rate in the city due to delayed home purchase. It is also likely that the impact of forced evictions works through the precautionary savings behavior of the home-owners, since they are likely to receive less compensation than the market value of their properties if being evicted. I find that both channels have some explanatory power explaining the increase in household savings rate due to forced evictions.

This chapter is related to two strands of literature. First, it links to the research establishing the link between housing and savings rate (Campbell and Cocco 2007; Chamon and Prasad 2010; Engelhardt 1996). Using micro data from U.K., Campbell and Cocco (2007) provide empirical evidence that household consumption increases as house price booms, especially for older homeowners. Chamon and Prasad (2010) find that owning a home is associated with sharply lower savings rate (4-7 percentage points) among young households in China. Going one step further, the empirical evidence in this paper suggests that owning a home with uncertainty of being demolished will induce households to save more. This is consistent with the empirical findings in Engelhardt (1996) that households experiencing real housing capital loss are likely to save more in order to offset such loss. In addition to the empirical evidences discussed above, the theoretical model proposed by Hu (2005) also suggests that a higher exogenous moving rate (such like a higher probability of forced move) should induce more precautionary savings.

Second, this paper provides supporting evidence on precautionary savings motive in China. Previous literature suggests that precautionary savings motive is a potential explanation to the increasing household savings rate in China (Meng 2003; Blanchard and Giavazzi 2006; Chamon and Prasad 2010; Giles and Yoo 2007; Feng, He and Sato 2011). Several factors, such as increases in education, health and housing expenditure, and uncertainty in pension wealth have been suggested to induce precautionary savings. However, there are no existing studies to examine precautionary savings motive due to home ownership uncertainty.

The primary contribution of this chapter is that it documents the rising home ownership insecurity problem in China and its macroeconomic consequences. Since there is no official statistics on forced evictions in China, the dataset collected from online news provides a great opportunity to unveil the distribution of forced eviction and to study its consequences. Second, this paper adds to the literature in understanding the causes of rising household savings rate in China. In align with the hypothesis on precautionary savings, this paper provides robust empirical evidences suggesting that home ownership uncertainty may explain some of the rising household savings rate in china.

The chapter is structured as follows. Section 4.2 provides the institutional background about increasing demolition in China; Section 4.3 introduces the hypothesis to be tested, data sources, and identification strategies; Section 4.4 discusses the main findings; Section 4.5 shows the possible channels; and Section 4.6 concludes.

4.2 Background of Home Demolition in China

Home demolition is an increasing concern in China accompanied with the rapid growth of infrastructure investment and urbanization process. According to the China Household Finance Survey (CHFS) conducted in year 2011, 8.7% of the surveyed urban households have experienced home demolition since year 2000, with an average of 129.15 square meters being demolished (see Table B.25).

The average monetary compensation is around 276.8 thousands yuan, which is more than 11 times of the average urban household income in 2011. [1] However, it seems that the households being demolished are not very satisfied with their compensation. Around half of the households are not satisfied (or very unsatisfied) with the amount being compensated. 4.73% of the households even received nothing for the demolition.

More households in eastern China (11.54%) have experienced home demolition than households in central and western China (5.03%), with significantly more areas being demolished. Since housing price is generally higher in eastern China, the average monetary compensation in the east is more than three times as of the compensation in central and western China. However, the share of households unsatisfied with the compensation is not significantly lower in the east. One possible reason is that the money being compensated cannot match the average housing price in both the east and the central and western China.[2]

When home owners cannot reach an agreement with the party initiating demolition, such as property developers, conflicts arise. Some households refused to move even the real estate projects began around them, which resulted in forced evictions. Such conflicts sometimes evolved into violent incidents and brought media attention.[3] In the past decade, the incidences of forced evictions have been increasing dramatically both at the intensive (more cases in the same area) and extensive (more areas being affected) margin. The number of news

[1]The average urban household income is 23,979 yuan in 2011, according to the China Statistical Yearbook 2012.

[2]According to the survey team, the compensation to housing value ratio is around 0.8 in the east, 0.87 and 0.89 in central and western China. (The full report is available at http://money.163.com/13/0325/09/8QQ8K7C7002534M5.html)

[3]For example, Wall Street Journal reported a mapping project about violent forced evictions in China: http://blogs.wsj.com/chinarealtime/2010/10/29/chinas-blood-stained-property-map/

about forced evictions by searching Baidu News Archive, the world's largest Chinese news search engine, rose from 307 in 2004 to 125,000 in 2012.

4.3 Hypothesis, Data Sources, and Identification Strategies

4.3.1 Hypothesis

This chapter tries to understand the macroeconomic consequences of forced evictions. Specifically, being exposed to more forced evictions in a city increases residents' expectation of being evicted, which lowers the expected housing wealth since the compensation for demolition (if there is any) is usually below market value. Therefore, the following hypothesis is formed: more forced evictions in a city reported through media subsequently increases urban households' precautionary savings motive in that city, in order to offset the expected housing capital loss. In this paper, I only focus on urban households since they have easy access to media. The rural households may not have access to newspapers and internet, thus unlikely respond to media reporting of forced evictions.

4.3.2 Data Sources

The study will be carried out at the prefecture city level due to the type of data available for use.

Forced evictions

One of the key variables in this study is the level of home ownership insecurity being perceived by the households for each city in different years, which is indexed by the number of news about forced evictions in a certain city being reported yearly searched from Baidu News Archive. [4]For example, the resulted number of news by searching news titles containing“forced eviction” and “Guangzhou” from January 1st, 2010 to December, 31st, 2010 is 1,280. Therefore, the index for home ownership insecurity in Guangzhou in 2010 is 1,280. The larger the value suggests the worse the level of home ownership insecurity in a city in a certain year.

Since no published statistics in China reveals the incidence of forced eviction, media coverage is the best available resort to learn about forced evictions from the residents' perspectives. However, there may be several concerns regarding the manner that forced eviction is defined. First, it is likely that some forced evictions are not reported due to media censorship in China. However, this is less of a concern since households respond only to information they are exposed to about forced eviction, if there is any causal responses. Most people collect information from local newspapers, television news and internet, which is covered by Baidu News Archive. Therefore, the forced evictions being censored, without known to the public, only affect the information collected by the witnesses and victims of such events, which should be a rather tiny group relative to the city population.

Second, the level of media censorship may vary by cities, which suggests that institutional differences in cities may affect the number of news being reported about forced eviction. For example, in more liberal cities in China, local

[4]There is a growing trend of collecting data from online resources in social science research, such as Google Trends (Choi and Varian 2012; Vosen and Schmidt 2011; Askitas and Zimmermann 2009).

government may not be very sensitive about releasing forced eviction news to the public, so most of the forced eviction events may have been reported in those areas. In contrast, a conservative local government may impose more censorship on forced eviction news, so the cases of forced evictions being reported may be much less than it actually is. Such institutional differences in cities may bias the number of news being collected about forced evictions in different cities, which may also affect household savings rate in some ways. For example, it is possible that more liberal cities generally have higher GDP and personal income, so people save more. In order to deal with such concerns, city level economic and demographic indicators, such as GDP, fixed asset investment and population are controlled in the model when studying the causal impact between forced evictions and household savings rate. It is rather unlikely that non-economic factors correlated with institutional differences, such as local government ideologies, will affect household savings rate. However, I try to capture such factors through controlling some institutional variables as well. Third, one event of forced eviction may correspond to more than one news titles from the Baidu News Archive. However, this works in favor of the question to be addressed in the paper since more media attention indicates more informed the population is. Therefore, the counts of news titles on forced eviction essentially combine the number of forced eviction cases and the attention of such events among the public.

In order to reach a robust conclusion that insecure home ownership induces household saving behavior, I use two alternative measures of forced evictions as robustness checks. First, I further refine the searching criteria of forced eviction news in Baidu News Archives. In addition to the two key words "forced eviction" and city name, I further exclude the key word "demolition" from the searching criteria since Baidu search engine takes "demolition" as a synonym for "forced eviction", thus include the news titles containing "demolition" in the returned results as well. Second, I normalize the forced eviction index by the total population of the city. Intuitively, the normalized index measures the probability of being evicted in the city.

Household savings rate

Another key variable used in this study is household savings rate in prefecture level cities. Household savings rate in the prefecture city level is defined as 1-average per capita consumption expenditure/average per capita disposable income. The consumption and income data is collected from CEIC database which only includes the urban households. The data is sourced from National Bureau of Statistics (China). Other city statistics, such as GDP, fixed asset investment and population, is collected from China Economic and Social Development Statistical Database,[5] and the data source is China Statistical Yearbooks in various years. Since Baidu News Archive does not contain information prior to November 2003, the years relevant to this study span from 2004 to 2011.

Descriptions of key variables

Figure A.13 and A.14 shows the map series of media reports about forced eviction in 248 prefecture level cities in China in 2004 and 2011. Since the consumption and income data is not available for most of the cities in western China (mainly minority autonomous regions) in the CEIC dataset, the forced eviction index is not collected for those cities which are shaded in grey in the maps. However, the sample is still representative for China since 248 out of 283

[5]The database is available at http://tongji.cnki.net/kns55/Dig/dig.aspx with institutional access.

prefecture level cities have been included. It is clearly shown from the maps that the incidences of forced evictions have been increasing dramatically both at the extensive and intensive margin in the past eight years. Only 56 out of 248 cities were reported about forced eviction in 2004, while 203 cities had at least one piece of news about forced eviction in 2011. Among the cities being reported for forced eviction, the average number of news rose from 3.14 in 2004 to 27.33 in 2011, which grew almost 9 times.

Table B.26 describes the source of data, unit of measurement, as well as summary statistics for several key variables. The average household savings rate is 28.3 percent from 2004 to 2011, with a standard deviation of 6.76. However, the variation of forced eviction across cities is much larger. The average number of news being reported is 7.54 for a prefecture level city on an annual basis, but the standard deviation is 44.47. Guangzhou reported 1,280 pieces of news on forced eviction in 2010, the largest among all the observations. The average city in China has a GDP of 105.6 billion yuan, fixed asset investment of 37.5 billion yuan and total population around 4.3 million. Figure A.15 shows the growth trend of urban household savings rate and reported forced evictions from 2004 to 2011, both of which are rising over years.

4.3.3 Identification Strategies

The dataset to be estimated is a panel data consisting of 248 prefecture level cities in China and eight years (2004-2011). Since household savings rate is path dependent, it is necessary to include lagged savings rate in year $t - p$ ($p > 0$) to explain the variation of savings rate in year t. Therefore, the traditional Ordinary Least Squares estimation with fixed effects can be specified as follows:

$$Savings_{i,t} = \alpha + \gamma_1 Savings_{i,t-1} + \cdots + \gamma_p Savings_{i,t-p} + \beta \ln FE_{i,t-1} + \delta X_{i,t-1} + \sigma_i + \epsilon_{i,t} \tag{4.1}$$

However, the Ordinary Least Squares estimation with city fixed effects leads to biased coefficients due to the correlation between the demeaned lagged dependent variable and the error term, especially in the "small T, large N" context (Nickell, 1981). Therefore, the empirical specification here applies the Difference GMM estimation, which is proposed by Arellano and Bond (1991). Specifically, Arellano and Bond (1991) take the first difference of both dependent and independent variables of a panel dataset to get rid of the time-invariant country (city) specific characteristics. In addition, the lagged differenced dependent variable (savings rate) of order p is included in the right-hand side of the equation if we assume the dependent variable follows an AR(p) process. The equation to be estimated by the Arellano-Bond estimator is specified as follows:

$$\Delta Savings_{i,t} = \alpha + \gamma_1 \Delta Savings_{i,t-1} + \cdots + \gamma_p \Delta Savings_{i,t-p} + \beta \Delta \ln FE_{i,t-1} + \delta \Delta X_{i,t-1} + \Delta \epsilon_{i,t} \tag{4.2}$$

where $Savings_{i,t}$ is the average urban household savings rate in city i, time t; $\ln FE_{i,t-1}$ is the log value of forced eviction index in city i, time t; $X_{i,t}$ is other city economic and demographic statistics which may affect savings rate, including GDP, fixed asset investment, population and foreign direct investment in log forms. Specifically, GDP and population control for the size of the economy. Fixed asset investment controls for the size of the real estate sector which is likely to be correlated with housing price, and thus savings rate. Ideally, housing price should be controlled in the savings equation. However, since there is no data available for historical housing prices in China, I use fixed asset investment to proxy for housing prices at the city level. Foreign direct investment controls for the institutional differences across cities. As mentioned in Section 4.3, institutional differences in cities, such as openness, may affect the number of news being reported about forced eviction. Therefore I include foreign direct investment to control for the potential differences in terms of openness of a city. Finally, $\epsilon_{i,t}$ is the error term.

Since $Savings_{i,t-p}$ is correlated with the error term $\Delta\epsilon_{i,t}$ only when p=1, the Arellano-Bond estimator uses $Savings_{i,t-p}$ $(p \geq 2)$ to instrument $\Delta Savings_{i,t-1}$ which eliminate the estimation bias due to the correlation between lagged dependent variable and the error term. β is the coefficient of interest, which indicates the impact of insecure home ownership in period $t-1$ on household savings rate in period t.

Two specification tests are to be carried out to test the validity of the models. First, the Sargan test statistics will be reported to test the identification restrictions, with the null hypothesis that the instruments are valid. Second, the first- and second-order serial correlation test statistics will be reported to test the hypothesis that the error term is not serially correlated. If the model is correctly specified, the first-order serial correlation test should reject the null hypothesis while the second-order serial correlation test should not reject the null.

If the dependent variable is highly persistent over time (similar to a random walk), then the lagged dependent variables become weak instruments for differenced lags. Thus the Difference GMM estimation suffers from finite sample bias. In order to deal with this problem, Arellano and Bover (1995) and Blundell and Bond (1998) propose the System GMM estimation that combines regressions both in differences and in levels into one system. However, the Difference GMM estimation is a preferred specification in this paper suggested by the specification tests. [6] Moreover, the dependent variable in our dataset does not contain unit root indicated from several unit root tests on panel data. [7]

4.4 Findings

Before diving into the Difference GMM estimation results, Table B.26 exploits the potential factors which affect the incidence of forced evictions. The dependent variable is the incidence of forced evictions in log form, while explanatory variables include lagged values of GDP, fixed asset investment, total population and foreign direct investment in log forms, as well as dummies for whether a city is provincial capital or municipality, year fixed effects and province fixed effects. The robust standard errors are clustered at the city level. Column 1-4 controls for the four lagged economic variables one at a time, and column 5 controls for all of the four economic variables and other explanatory variables mentioned above. It can be seen from the first four columns that higher GDP and population, more fixed asset investment and foreign direct investment are

[6] Table B.32 provides the results for System GMM estimation. However, 5 out of the 15 models cannot pass the Sargan Test; none of the models pass the first order serial correlation test; and 5 models cannot pass the second order serial correlation test. Therefore the System GMM models are likely to be mis-specified.

[7] The Levin-Lin-Chu unit-root test (Levin et al., 2002) and Harris-Tzavalis unit-root test (Harris and Tzavalis, 1999) are two tests with null hypothesis that "panels contain unit roots". The test statistics in both tests reject the null hypothesis at the 0.01 level. The Hadri LM test (Hadri, 2000) is based on the null hypothesis that "all panels are stationary" and the alternative hypothesis as "some panels contain unit roots". The test statistic cannot reject the null hypothesis at the 0.05 level. Therefore, all the evidences suggest that the savings data does not contain unit root.

all significantly positive correlated with more forced evictions in the next year in a certain city. However, when the four variables are controlled at the same time, GDP is the only dominant factor significantly affecting forced evictions, while other three variables are no longer significant. In addition, provincial capitals and municipalities are significantly associated with more forced evictions comparing to other cities.

Table B.27-B.29 report the impact of reported forced evictions on urban household savings rate using three different forced eviction measures. All regressions are estimated using the Arellano-Bond Estimator following Equation (4.1). The covariates enter into the model with different assumptions in Panel A and Panel B. Panel A assumes all the right hand side variables to be strictly exogenous, while Panel B relaxes the restrictions by assuming all the right hand side variables to be endogenous. For each panel, there are five specifications including different sets of control variables. In addition to the coefficients on forced evictions, the number of instruments and the statistics on the two specification tests are also reported in the tables.

Table B.27 reports the estimated coefficient of interest, β, using the first measure of reported forced eviction, i.e., the logged value of returned number of news by searching keywords "forced eviction" and the city name within a certain year. The coefficients in all the five specifications in Panel A are positive and significant at the 0.05 level with a magnitude between 0.39 and 0.41. This suggests that worse insecurity of home ownership, as indicated by higher forced eviction index, leads to higher household savings rate at the prefecture city level in the subsequent year. Panel B reports more conservative estimation assuming all the right-hand-side variables to be endogenous. Three out of the five coefficients are positive and significant at the 0.10 level with a magnitude between 0.67 and 0.88. More specifically, if the forced eviction index is increased by the average annual growth rate suggested from the data, 161%, urban household savings rate is expected to increase by 0.64 to 0.84 percent. In addition, the null hypothesis of the Sargan test cannot be rejected for all the ten specifications in Table B.27, supporting the statement that the specifications in the dynamic panel data model are not overidentified. Similarly, the serial correlation test suggests that the lagged error terms are not correlated, which satisfies the assumption of the Arellano-Bond estimator. [8]

Table B.28 presents the results using the second measure of reported forced eviction, which filters out the key word "demolition" from the first measure. Again, all the coefficients on reported forced evictions are positive. However, the number of significant coefficients (at the 0.05 level) decreases, especially for the model assuming strict exogeneity for all the covariates. Only one of the coefficients in Panel A is significant at the 0.10 level. The other four coefficients are marginally significant though. The estimated coefficients in Panel B are similar to the counterparts in Table B.27. Again, all the specification tests in Table 5 support the validity of the model.

Table B.29 shows the estimation results using the third way of measuring reported forced eviction, which is the first measure normalized by population of the city, representing the probability of being evicted in the city. All the five coefficients in Panel A are positive and significant at the 0.01 level, which is consistent with the findings in Table B.27 and B.28. However, all the coefficients lose their significance after applying more conservative estimation in Panel B,

[8] If the lagged error terms are not correlated, the first order serial correlation test should always be rejected, and the second order serial correlation test should not be rejected.

which indicates that the estimation using normalized reported forced evictions is not robust across different model settings. Such fact reflects the possibility that households are more responsive to the absolute number of news being reported instead of taking care of the population size in the city.

To summarize, the estimation results in Table B.27-B.29 supports the hypothesis that more forced evictions in a city reported through media subsequently increases urban households' precautionary savings motive in that city. The results are robust to different measures of forced evictions at the absolute value, and are consistent across various specifications.

4.5 Possible Mechanisms

As mentioned in the introduction section, there are two possible mechanisms which may lead to the impact of forced evictions on household savings rate. First, it is likely that the impact of forced evictions on household savings rate works directly through a reduction in home sales. If forced evictions discourage home sales, it may increase the average household savings rate in the city due to delayed home purchase. It is also likely that the impact of forced evictions works through the precautionary savings behavior of the home-owners, since they are likely to receive less compensation than the market value of their properties if being evicted.

I test the first channel by estimating the impact of forced evictions on residential floor area sold in the cities using the same difference GMM specification. As suggested in Table B.30, I find that increases in forced evictions significantly discouraged home sales. Among the ten coefficients on forced evictions using the first two measures of forced eviction indices, seven of them are significantly negative, which provides supporting evidence to the first channel.

In order to test the second channel related to precautionary savings behavior, I estimate the impact of reported forced evictions using two restricted samples where compensation for evictions is likely to be lower, since households may worry more about the devaluation of housing assets thus engage more in precautionary savings in cities being compensated less from forced evictions. Therefore, in Table B.31, Panel A, I restrict the sample to cities in central and western China. According to the CHFS survey (Table B.24), the compensation about home demolition is far less in central and western China than in eastern China. Therefore, the magnitude of forced eviction's impact on savings is likely to be larger in central and western China compared to the full sample. The magnitude of the estimated coefficients on three measures of forced evictions is consistently larger than the corresponding estimates in Panel B of Table B.27-B.29, which works in favor with the above hypothesis.

In some major cities in China, such as provincial capitals, government may be willing to pay higher compensation to the evicted households to avoid possible social unrest due to forced evictions. In this regard, forced evictions may not lead to precautionary savings behaviors. Therefore, I exclude all the provincial capitals from the sample in Panel B. Again, the magnitudes of all the estimated coefficients for the first two measures of forced evictions are generally larger than the coefficients estimated from the full sample. The coefficients on the normalized measure of forced evictions are not significant, which is consistent with the estimated results in Panel B of Table B.29. Overall, the robustness checks in Table B.31 suggest that the precautionary savings motive driven by reported forced evictions is more pronounced in cities where the compensation for evictions is less.

4.6 Conclusion

Using a unique dataset collected from online news coverage, I provide some suggestive evidences about the causation between rising home insecurity, as measured by the reported incidence of forced eviction, and rising urban household savings rate in China. By applying dynamic panel data models, I find that worse insecurity of home ownership, as indicated by more frequent forced evictions, leads to higher urban household savings rate at the prefecture city level in the following year. This impact is likely to be attributed to the discouraged home sale due to forced evictions, and the precautionary savings behavior of home owners due to less compensation from forced evictions.

The empirical findings in this chapter imply that home ownership uncertainty plays a role in explaining the rising savings rate in China, which is consistent with the precautionary savings motive. Therefore, improved enforcement on home ownership protection not only reduces social conflicts due to forced evictions, but also has its macroeconomic implications. Since China stresses boosting consumption as one of its long term economic growth strategies, the government should reduce the barriers preventing people from spending their money, such like insecure property ownership as discussed in this chapter.

APPENDIX A

FIGURES

Figure A.1: The Decreasing Trend of Passenger Train Stations (1996-2009)

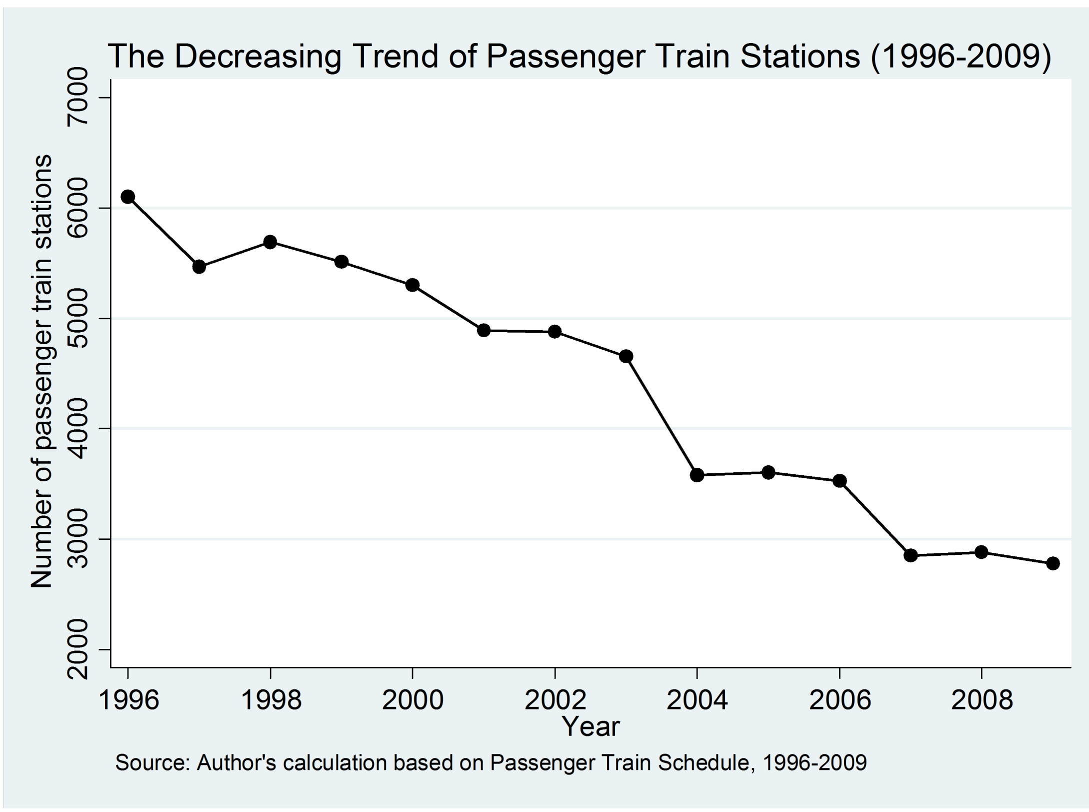
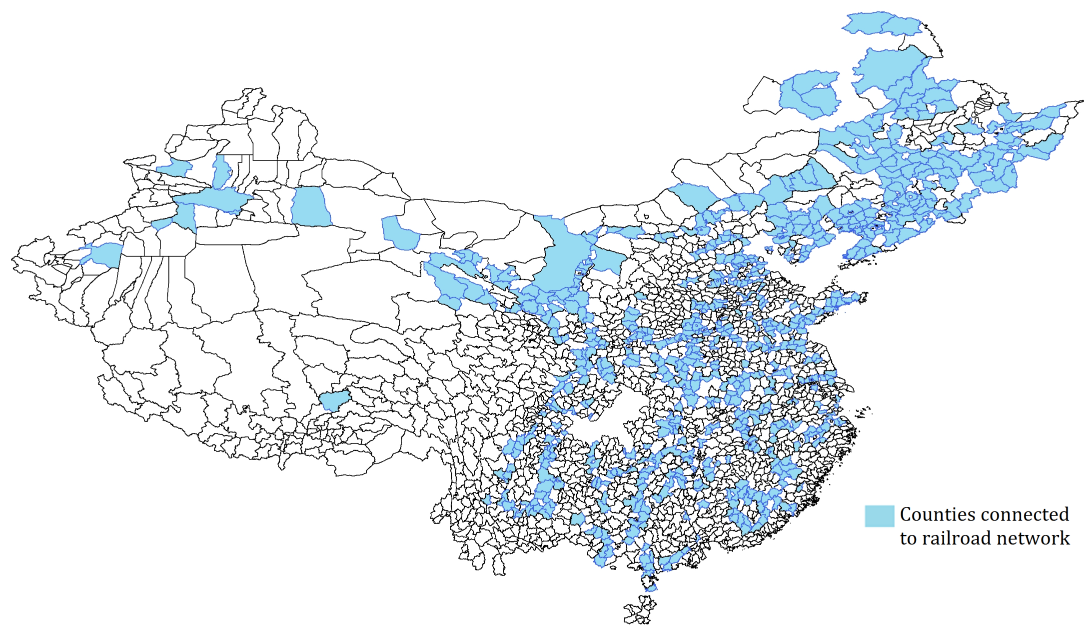
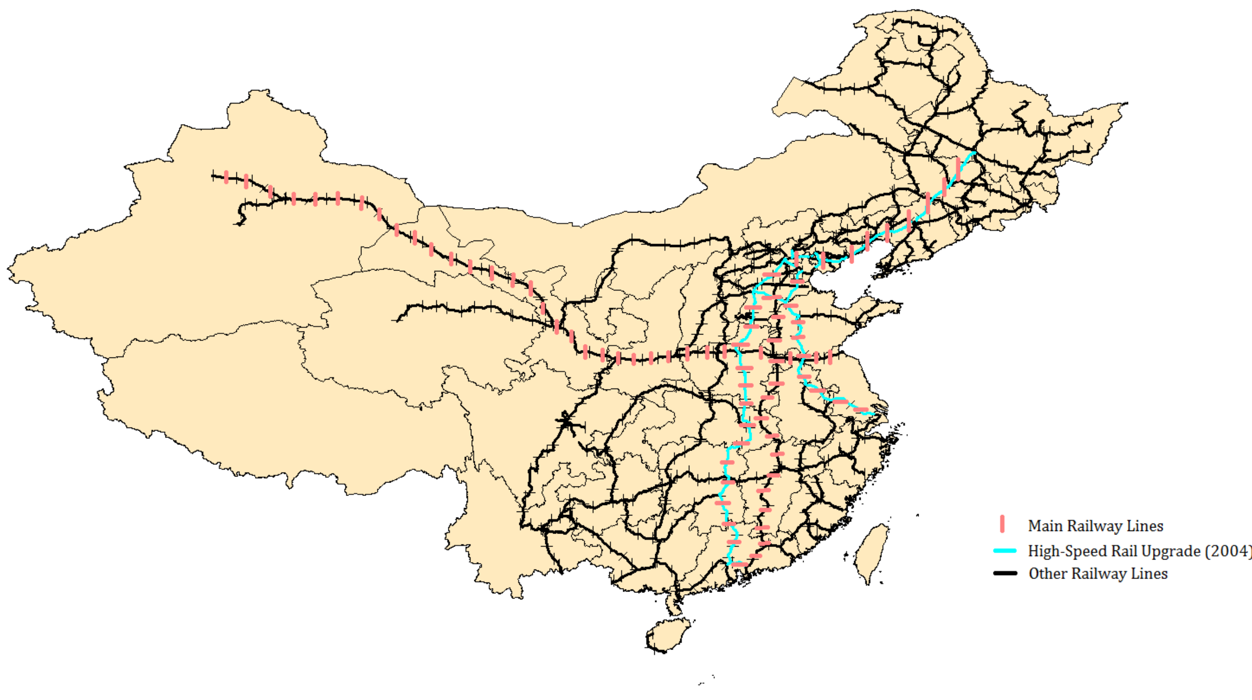


Figure A.2: Counties Connected to Railroad Network by Year 2007



GIS source: China Data Center (University of Michigan) and *People's Republic of China Railroad Atlas*. Blank areas are urban districts of prefecture-level cities, which are not included in our analysis.

Figure A.3: High-Speed Rail Upgrade in 2004



GIS source: China Data Center (University of Michigan).

Figure A.4: High-Speed Rail Upgrade in 2007

GIS source: China Data Center (University of Michigan).

Figure A.5: Railway Accessibility and Railway Length (1996-2009)

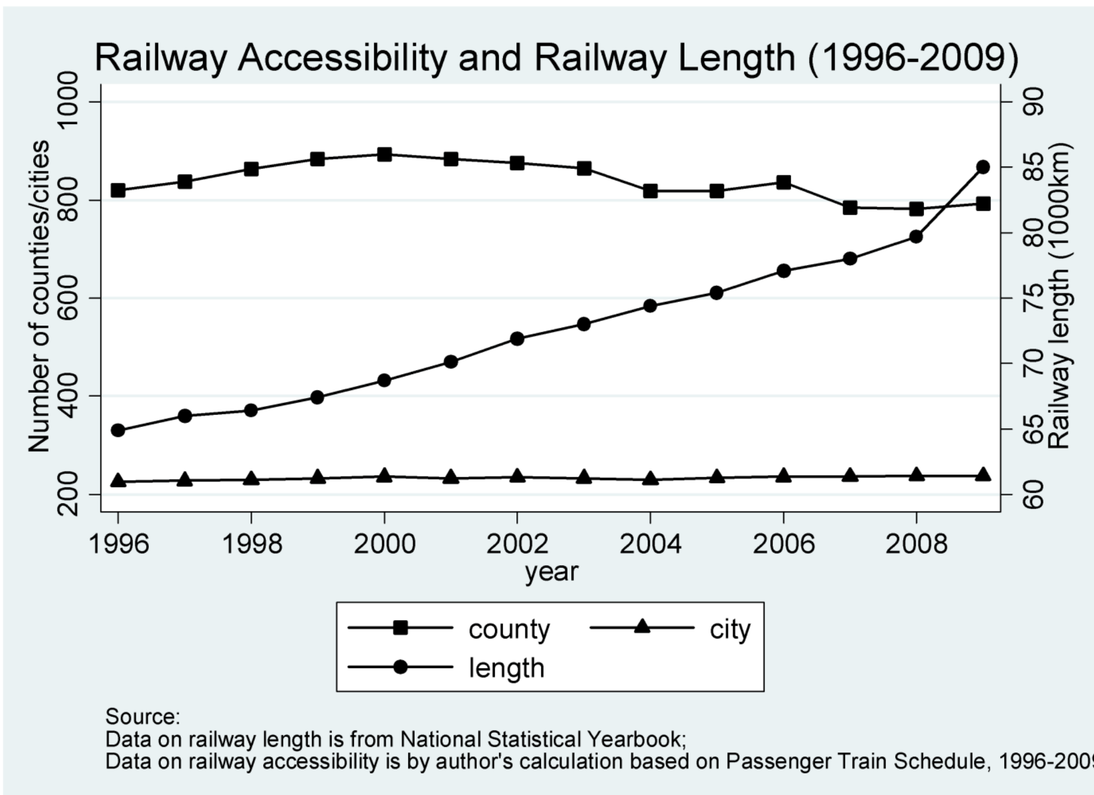


Figure A.6: Average Daily Train Stops by City and County (1996-2009)

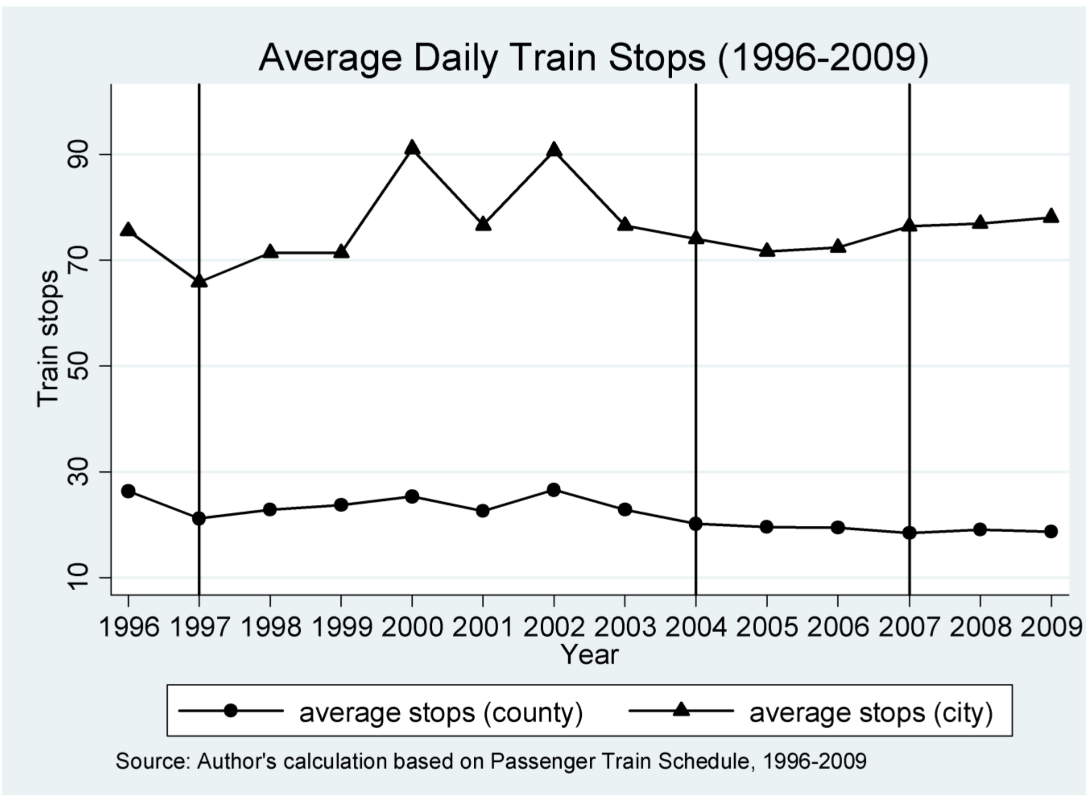
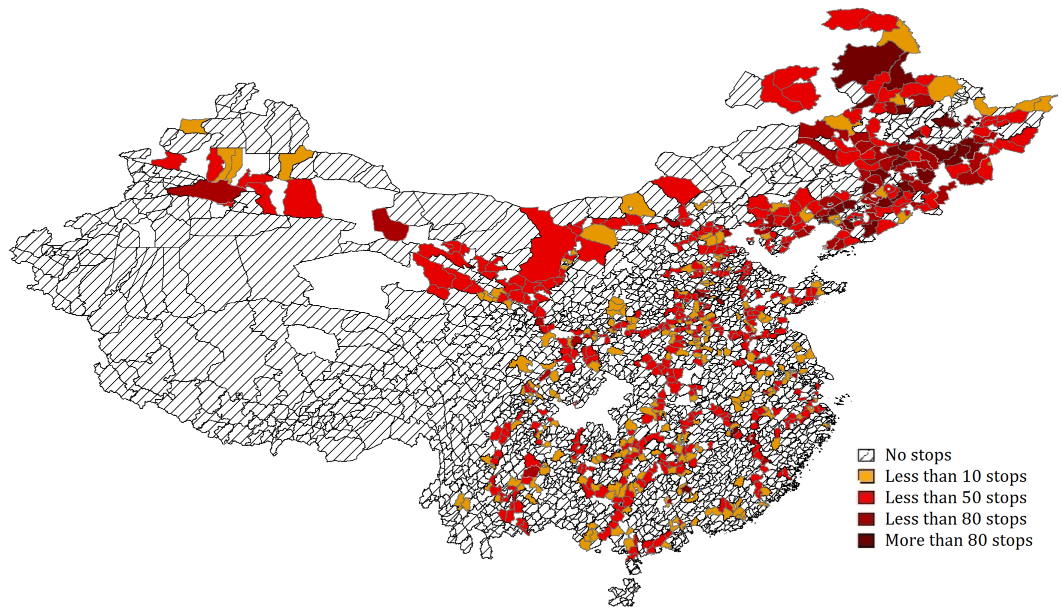
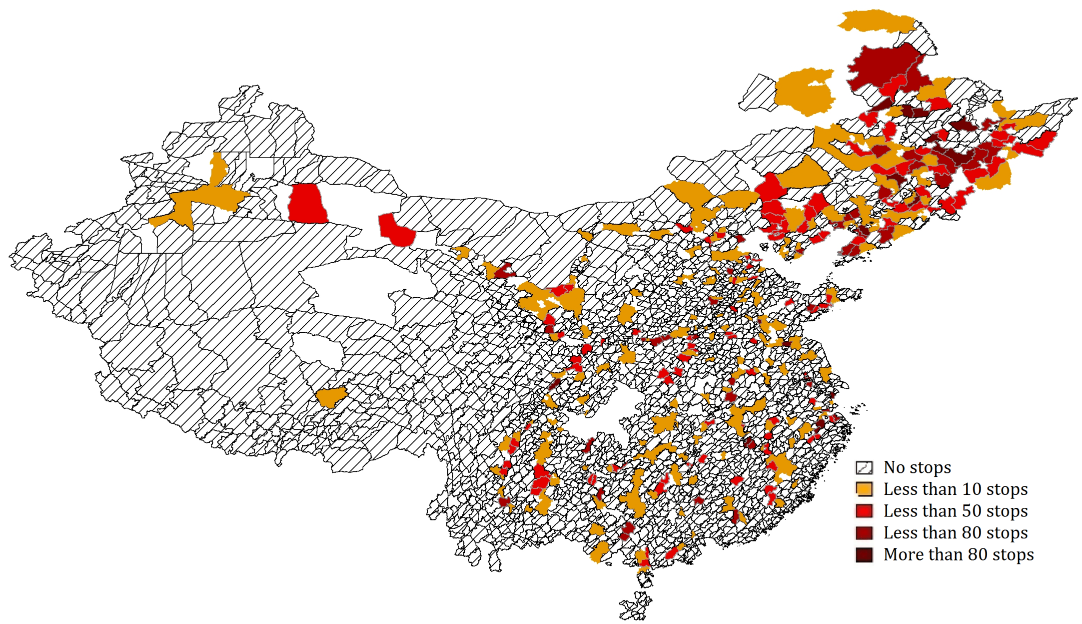


Figure A.7: Daily Average Train Stops in the Counties, 1996



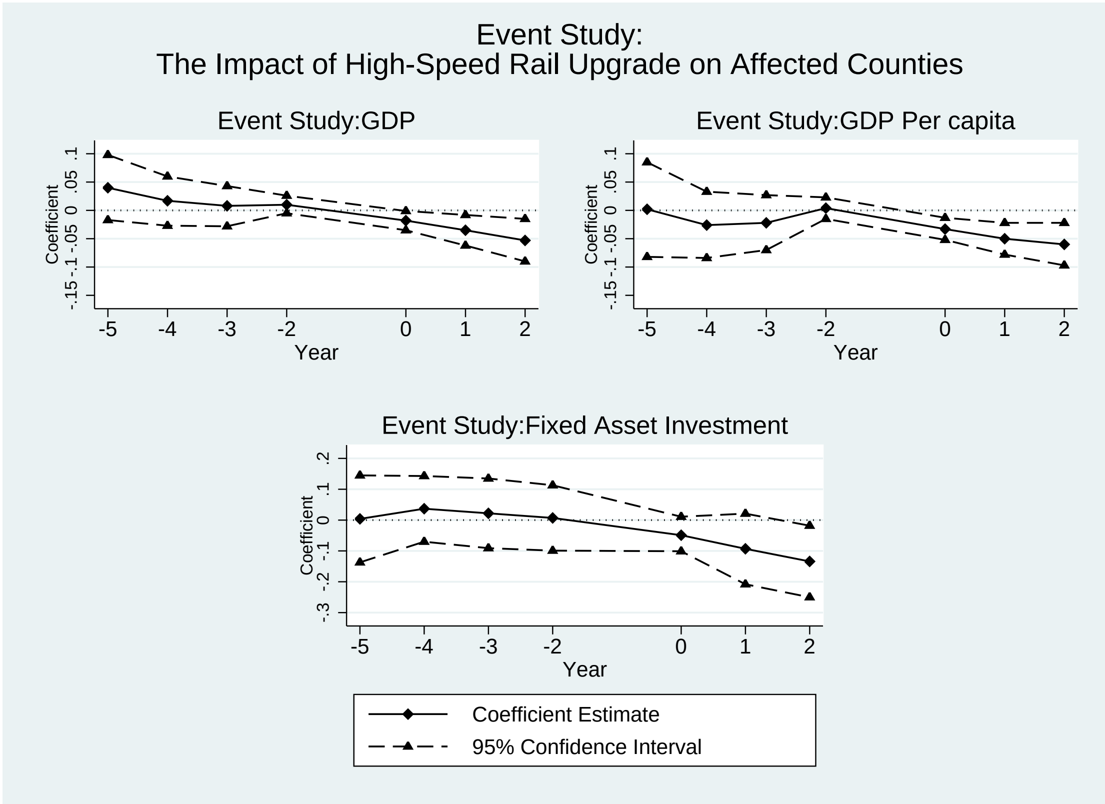
GIS source: China Data Center (University of Michigan). Blank areas are urban districts of prefecture-level cities, which are not included in our analysis.

Figure A.8: Daily Average Train Stops in the Counties, 2007



GIS source: China Data Center (University of Michigan). Blank areas are urban districts of prefecture-level cities, which are not included in our analysis.

Figure A.9: Event Study: The Impact of High-Speed Rail Upgrade on Affected Counties

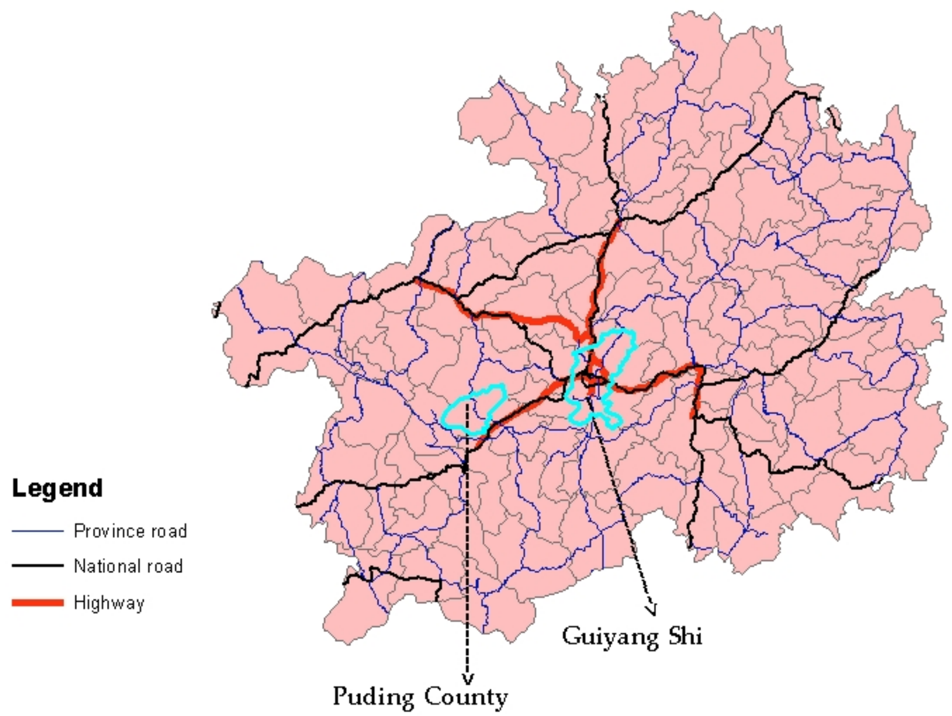


Notes: 1. Year 0 indicates year 2007, when the second round of high-speed rail upgrade was implemented. Year -1 (year 2006) is the baseline year for comparison. 2. For each coefficient, the 95% Confidence Interval is reported.

Figure A.10: Map of China

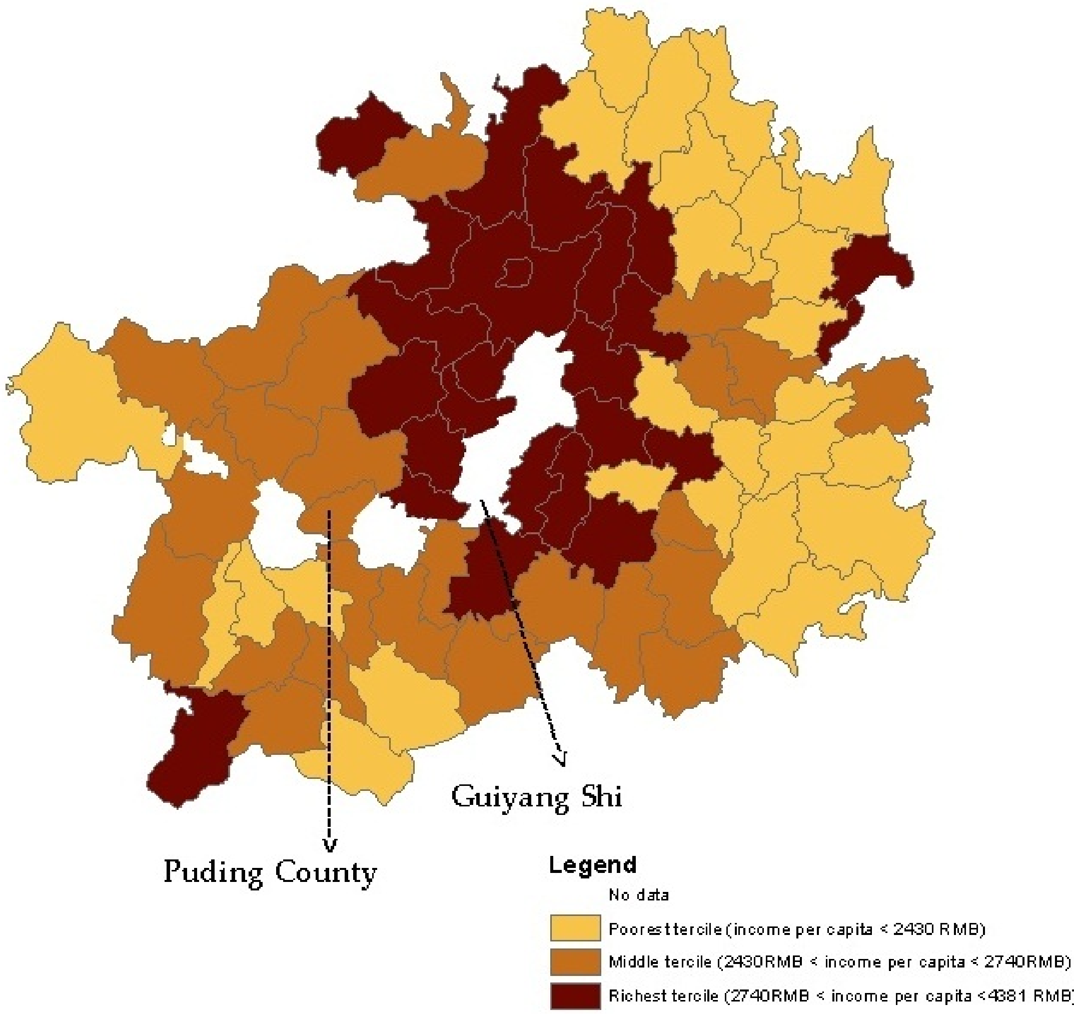
Source: China Data Center (University of Michigan).

Figure A.11: Map of Guizhou Province with road network



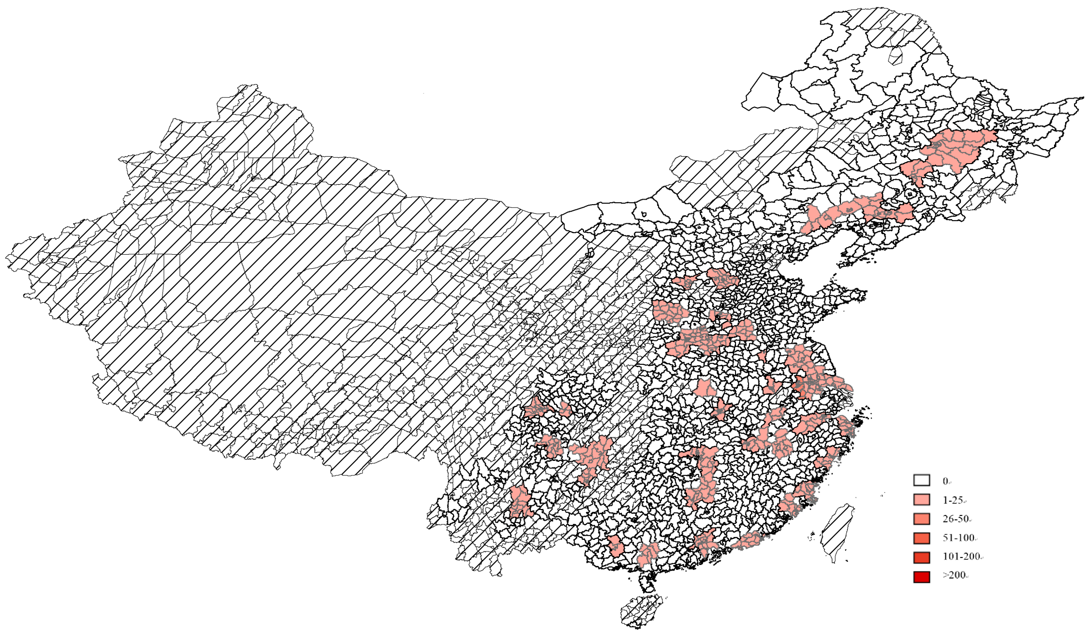
Source: China Data Center (University of Michigan).

Figure A.12: Income per capita of counties in Guizhou (year 2008)



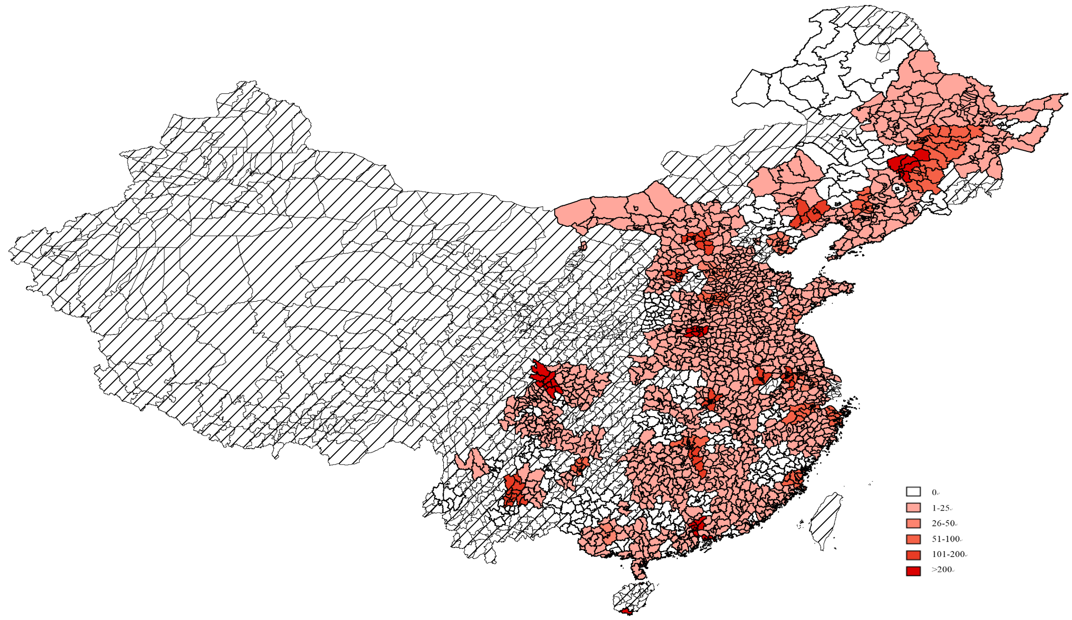
Source: China Data Center (University of Michigan).

Figure A.13: Reported Forced Evictions in China in 2004



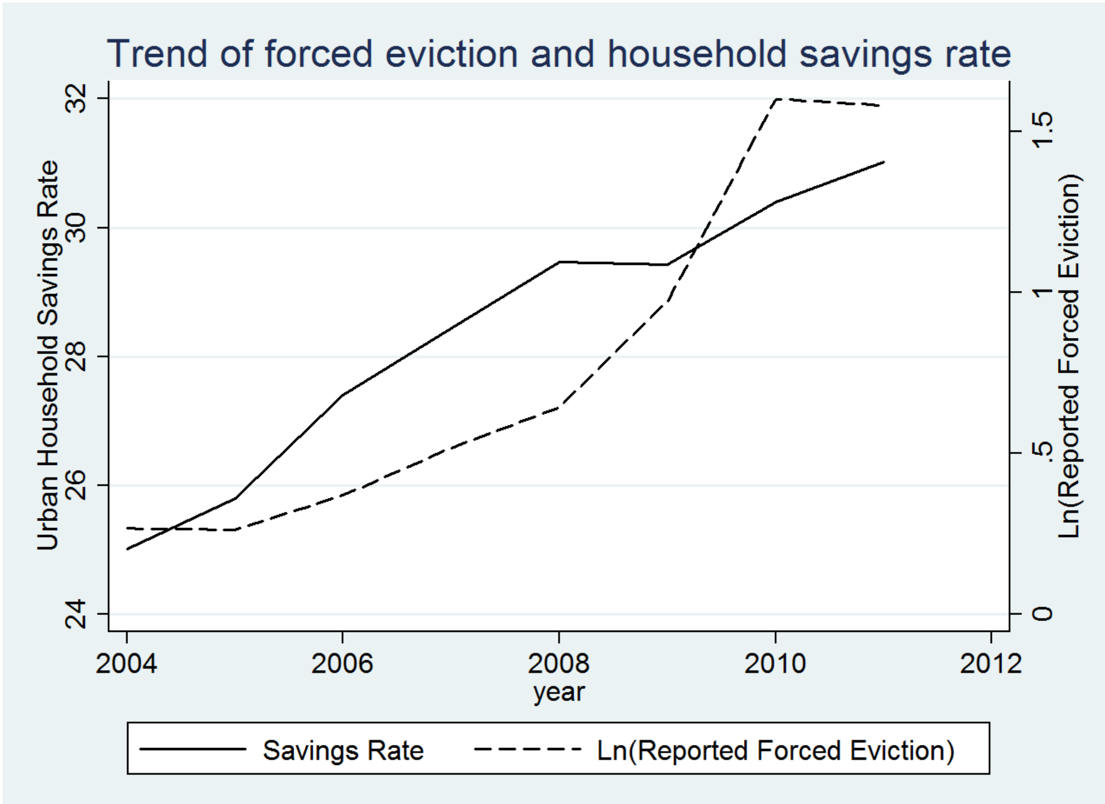
Note: Forced eviction index is defined as the number of news titles including key words “forced eviction” + “city name” in Baidu News Archives each year. Grey areas are the cities without statistical data on savings rate.

Figure A.14: Reported Forced Evictions in China in 2011



Note: Forced eviction index is defined as the number of news titles including key words "forced eviction" + "city name" in Baidu News Archives each year. Grey areas are the cities without statistical data on savings rate.

Figure A.15: Trend of Forced Eviction and Household Savings Rate:2004-2011



Note: Forced eviction index is defined as the number of news titles including key words “forced eviction” + “city name” in Baidu News Archives each year. Urban household savings rate is defined as 1-expenditure/disposable income, collected from CEIC database.

APPENDIX B

TABLES

Table B.1: Descriptive Statistics

	2003		2007		Source
A. Railway Status	Main RL	Other RL	Main RL	Other RL	
Number of Counties	171	786	171	786	*People's Republic of China Railroad Atlas*
Average Daily Train Services (with Stops)	27.67	21.74	20.79	17.62	*China Passenger Train Schedule (annually)*
	(26.22)	(22.91)	(20.04)	(17.87)	
B. Economic Outcomes					
GDP (100 million *yuan*)	47.19	36.72	94.77	73.61	*China Economic and Social Development Statistical Database*
	(44.29)	(37.59)	(107.65)	(74.94)	
GDP Per Capita (1000 *yuan*)	9.58	7.24	16.06	15.01	*China Economic and Social Development Statistical Database*
	(14.87)	(4.96)	(15.42)	(12.19)	
Fixed Asset Investment (100 million *yuan*)	15.03	12.56	41.87	36.87	*China Economic and Social Development Statistical Database*
	(17.16)	(15.82)	(40.59)	(37.30)	

Notes. 1. Main RL stands for "Main Railway Lines"; Other RL stands for "Other Railway Lines." 2. Mean and standard deviation (in parentheses) is reported for each of the variables.

Table B.2: The Impact of High-Speed Rail on County Economic Outcomes (OLS)

Table 2. The Impact of High-Speed Rail on County Economic Outcomes (OLS)

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.04		-0.08		-0.07
		(0.03)		(0.07)		(0.05)
HSR07*After	-0.04***	-0.05***	-0.05***	-0.05***	-0.10**	-0.11***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.04)	(0.04)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.98	0.92	0.91	0.89	0.90
Observations	4,689	7,498	4,614	6,431	3,953	6,327

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.3: The Impact of High-Speed Rail on County Economic Outcomes (Reduced Form)

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
Mainline	-0.47***	-0.15***	0.15***	0.61***	-0.36***	-0.36***
	(0.02)	(0.04)	(0.02)	(0.04)	(0.07)	(0.06)
Mainline*Year03		-0.01		0.04		0.05
		(0.01)		(0.06)		(0.04)
Mainline*Year04		-0.04**		0.00		0.00
		(0.02)		(0.03)		(0.05)
Mainline*Year05		-0.03		0.01		-0.01
		(0.02)		(0.04)		(0.06)
Mainline*Year06	-0.02**	-0.06**	-0.01	0.00	-0.05	-0.06
	(0.01)	(0.03)	(0.01)	(0.04)	(0.05)	(0.05)
Mainline*Year07	-0.04***	-0.07***	-0.04***	-0.03	-0.08	-0.09
	(0.01)	(0.03)	(0.01)	(0.04)	(0.05)	(0.06)
Mainline*Year08	-0.05***	-0.08***	-0.07***	-0.06	-0.13**	-0.14**
	(0.02)	(0.03)	(0.02)	(0.04)	(0.06)	(0.07)
Mainline*Year09	-0.07***	-0.10***	-0.08***	-0.07	-0.16**	-0.17***
	(0.02)	(0.03)	(0.02)	(0.04)	(0.07)	(0.06)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.98	0.92	0.91	0.89	0.90
Observations	4,689	7,498	4,614	6,431	3,953	6,327

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.4: The Impact of High-Speed Rail on County Economic Outcomes (2SLS)

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.08**		-0.06		-0.10
		(0.03)		(0.09)		(0.07)
HSR07*After	-0.04***	-0.04***	-0.06***	-0.05***	-0.10***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.98	0.92	0.91	0.89	0.90
Observations	4,689	7,498	4,614	6,431	3,953	6,327

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level. 3. The first stage F statistic is 12.16 for column 1, 3, 5; 71.91 and 12.16 for the two endogenous policy variables in column 2, 4, 6.

Table B.5: The Impact of High-Speed Rail on County Economic Outcomes (OLS, Collapsed Data)

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.04		-0.09		-0.09*
		(0.03)		(0.09)		(0.05)
HSR07*After	-0.04**	-0.05**	-0.05***	-0.05***	-0.11**	-0.12**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.98	0.95	0.92	0.89	0.92
Observations	1,880	2,819	1,880	2,528	1,616	2,458

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. Year fixed effect instead of year*province fixed effect is used in the regressions since the estimation of year by province trend requires a panel data of more than two periods. 3. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.6: The Impact of High-Speed Rail on County Economic Outcomes (2SLS, Collapsed Data)

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.08**		0.22***		0.11
		(0.03)		(0.07)		(0.09)
HSR07*After	-0.04***	-0.04***	-0.06***	-0.07***	-0.11***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.98	0.95	0.63	0.89	0.72
Observations	1,880	2,819	1,880	2,528	1,616	2,458

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. Year fixed effect instead of year*province fixed effect is used in the regressions since the estimation of year by province trend requires a panel data of more than two periods. 3. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.7: The Impact of High-Speed Rail on Prefecture Level City Economic Outcomes (OLS)

Panel A: Log Regressions						
	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		0.01		0.03		0.02
		(0.03)		(0.02)		(0.06)
HSR07*After	-0.01	0.01	-0.01	0.02	0.04	0.01
	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.05)
City Fixed Effect	Yes	Yes	Yes	Yes		Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes		Yes
R-Squared	0.99	0.99	0.98	0.98	0.97	0.96
Observations	1,176	1,884	1,185	1,896	1,177	1,883

Panel B: Level Regressions						
	Dependent Variables					
	GDP (100 million)		GDP Per Capita (yuan)		Fixed Asset Investment (100 million)	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		122.82		1141.43		87.23**
		(100.77)		(1153.80)		(43.24)
HSR07*After	167.61**	216.69**	445.97	461.56	100.83**	109.12**
	(75.81)	(106.42)	(954.34)	(1211.27)	(41.41)	(47.20)
City Fixed Effect	Yes	Yes	Yes	Yes		Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes		Yes
R-Squared	0.95	0.99	0.95	0.91	0.90	0.80
Observations	1,176	1,884	1,185	1,896	1,177	1,883

Notes . 1. Robust standard errors clustered at city level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.8: Heterogeneous Impacts of High-Speed Rail Upgrade in Different Sectors

	Panel A: OLS Estimation			
	Dependent Variables			
	Ln (Industrial Sector Value Added)		Ln (Service Sector Value Added)	
	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.05		-0.05
		(0.05)		(0.03)
HSR07*After	-0.03	-0.04	-0.03*	-0.03
	(0.02)	(0.03)	(0.02)	(0.02)
County Fixed Effect	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes
R-Squared	0.99	0.96	0.98	0.96
Observations	4,705	7,528	3,564	5,819
	Panel B: IV Estimation			
	Dependent Variables			
	Ln (Industrial Sector Value Added)		Ln (Service Sector Value Added)	
	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.09		-0.02
		(0.05)		(0.04)
HSR07*After	-0.03	-0.03	-0.03**	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
County Fixed Effect	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes
R-Squared	0.98	0.96	0.98	0.96
Observations	4,705	7,528	3,564	5,819

Notes. 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level. 3. LIML instead of 2SLS estimator is used in Column 4 of Panel B due to asymmetric and singular variance matrix problem.

Table B.9: The Impact of High-Speed Rail Upgrade Interacted with Distance to High-Speed Train Station

	Dependent Variables		
	Ln GDP	Ln GDP Per Capita	Ln Fixed Asset Investment
HSR07*After*Distance	0.00	0.00	-0.07
	(0.01)	(0.01)	(0.05)
HSR07*After*Distance Squared	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
HSR07*After	-0.04**	-0.05**	-0.06
	(0.02)	(0.02)	(0.06)
Distance	-0.42***	-0.04***	-0.32***
	(0.00)	(0.00)	(0.04)
Distance Squared	0.02***	0.00***	0.01***
	(0.00)	(0.00)	(0.00)
Distance*After	-0.01	0.00	0.01
	(0.01)	(0.01)	(0.02)
Distance Squared*After	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Distance*HSR07	0.50***	0.08***	0.08*
	(0.01)	(0.01)	(0.05)
Dsitance Squared*HSR07	-0.02***	-0.00***	0.00
	(0.00)	(0.00)	(0.00)
County Fixed Effect	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes
R-Squared	0.99	0.91	0.88
Observations	4,054	3,979	3,374

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level. 3. The unit of distance is 100 kilometers.

Table B.10: Channels: Increased Trade Cost in Affected Counties (2005-2009)

	Dependent variables		
	Ln GDP	Ln GDP Per Capita	Ln Fixed Asset Investment
HSR07*After	-0.05***	-0.05***	-0.12***
	(0.02)	(0.02)	(0.04)
Train service not reduced	0.09***	0.63***	-1.29***
	(0.02)	(0.02)	(0.06)
HSR07*After*Train service not reduced	0.02	0.00	0.04
	(0.02)	(0.02)	(0.05)
County Fixed Effect	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes
R-Squared	0.99	0.92	0.89
Observations	4,689	4,614	3,953

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.11: Channels: Diverted Economic Activities to Large Cities (2005-2009)

	Dependent variables		
	Ln GDP	Ln GDP Per Capita	Ln Fixed Asset Investment
HSR07*After	-0.09***	-0.08***	-0.12**
	(0.03)	(0.03)	(0.05)
Connected to Highway before 2007	0.84***	0.81***	-0.43***
	(0.01)	(0.00)	(0.05)
HSR07*After*Connected to Highway before 2007	0.06*	0.03	0.01
	(0.03)	(0.03)	(0.06)
County Fixed Effect	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes
R-Squared	0.99	0.92	0.89
Observations	4,689	4,614	3,953

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.12: The Impact of Speed Acceleration on County Economic Development, 1996-2003

	Dependent Variables					
	Ln GDP		Ln GDP Per Capita		Ln Fixed Asset Investment	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Speed97	0.02	-0.06	0.09	0.20*	0.21	0.10
	(0.03)	(0.06)	(0.06)	(0.11)	(0.15)	(0.22)
Speed98	0.09*	0.07	0.18*	0.19*	-0.31*	-0.33
	(0.05)	(0.08)	(0.10)	(0.10)	(0.18)	(0.26)
Speed00	-0.02	0.02	-0.07	-0.08	0.04	-0.02
	(0.03)	(0.04)	(0.05)	(0.06)	(0.11)	(0.24)
Speed01	-0.01	0.01	-0.06	-0.06	-0.07	-0.08
	(0.02)	(0.01)	(0.07)	(0.06)	(0.09)	(0.07)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.99	0.94	0.94	0.83	0.83
Observations	4,079	4,079	3,008	3,008	3,070	3,070

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level. 3. First stage F statistics are 29.5, 48.71, 108.86 and 8775.15 for the four endogenous variables, respectively.

Table B.13: The Impact of High-Speed Rail on Demographics (OLS)

	Dependent Variables							
	Ln Total Population		Ln Rural Population		Ln Total Households		Ln Rural Households	
	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009	2005-2009	2002-2009
HSR04*After		-0.01		-0.01		-0.01		-0.02
		(0.00)		(0.03)		(0.01)		(0.03)
HSR07*After	0.00	0.01**	0.01	0.01	0.00	0.00	0.00	0.01
	(0.00)	(0.00)	(0.03)	(0.02)	(0.00)	(0.01)	(0.04)	(0.02)
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.99	0.99	0.95	0.96	0.99	0.99	0.94	0.96
Observations	4,792	7,696	4,672	7,559	4,791	7,695	4,678	7,573

Notes . 1. Robust standard errors clustered at county level are reported in parentheses. 2. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.14: The Impact of High-Speed Rail Upgrade in Different Service Industries

	Dependent Variables: Ln (Employment) in Different Industries			
	Hotel and Restaurant	Financial Services	Real Estate	Rental Services
HSR(04 or 07)*After	-0.01	0.09*	-0.01	-0.30
	(0.06)	(0.05)	(0.14)	(0.22)
County Fixed Effect	Yes	Yes	Yes	Yes
Province*Year Fixed Effect	Yes	Yes	Yes	Yes
R-Squared	0.86	0.86	0.71	0.76
Observations	1,968	1,972	1,861	1,529

Notes . 1. Dependent variables are county level logged employment in different service industries collected from 2000 and 2010 China Population Census. 2. Robust standard errors clustered at county level are reported in parentheses. 3. * significant at the 0.1 level; ** significant at the 0.05 level; *** significant at the 0.01 level.

Table B.15: Road Access in Three Surveyed Villages

	2004	2006	2009	2011
Administrative Village I				
Natural Village 1	0	0	0	0
Natural Village 2	1	1	1	1
Natural Village 3	0	0	0	0
Natural Village 4	0	1	1	1
Natural Village 5	0	0	1	1
Natural Village 6	0	0	1	1
Natural Village 7	1	1	1	1
Natural Village 8	0	0	1	1
Natural Village 9	1	1	1	1
Administrative Village II				
Natural Village 1	0	0	0	0
Natural Village 2	1	1	1	1
Natural Village 3	0	0	0	0
Natural Village 4	0	1	1	1
Administrative Village III				
Natural Village 1	1	1	1	1
Natural Village 2	1	1	1	1
Natural Village 3	1	1	1	1
Natural Village 4	1	1	1	1

Source: Authors' Survey (2005, 2007, 2010, 2012).
Note: "1" denotes there is road access in the village in the specific year; "0" otherwise.

Table B.16: Agricultural Income Sources in Three Surveyed Villages

First Wave: Year 2004 (N = 758)			
Income sources (%)	Mean	Std	Share > 0
Corn	46.24	22.03	98.28
Paddy	11.87	17.68	46.04
Rapeseed	20.89	15.41	88.26
Vegetable	3.36	9.08	33.91
Fruit	2.20	7.02	20.71
Poultry	2.43	8.98	25.99
Livestock	11.93	21.21	29.95
Forestry	0.23	3.81	1.45
Fishing	0.84	6.46	2.11

Third Wave: Year 2009 (N = 759)			
Income sources (%)	Share	Std	Share > 0
Corn	42.04	24.34	94.99
Paddy	10.00	17.76	43.87
Rapeseed	16.00	12.40	93.15
Bean	4.00	7.62	82.35
Fruit	4.00	13.23	29.78
Poultry	2.00	10.60	37.94
Livestock	18.00	27.91	30.30
Forestry	0.00	5.00	5.27
Fishing	1.00	9.68	3.03

Second Wave: Year 2006 (N = 765)			
Income sources (%)	Share	Std	Share > 0
Corn	38.25	21.55	94.90
Paddy	11.98	17.18	47.32
Rapeseed	19.08	14.28	92.81
Other grain	1.99	4.50	33.33
Vegetable	11.20	14.57	85.49
Fruit	3.34	10.61	23.27
Poultry	1.56	6.42	20.92
Livestock	10.21	19.24	27.32
Forestry	0.32	2.68	2.88
Fishing	0.02	11.20	3.14

Fourth Wave: Year 2011 (N = 763)			
Income sources (%)	Share	Std	Share > 0
Corn	37.17	24.81	94.50
Paddy	8.01	16.31	30.93
Rapeseed	20.79	17.24	86.89
Bean	5.94	8.99	69.07
Vegetable	7.84	15.79	87.16
Fruit	5.38	16.10	16.12
Poultry	2.29	8.60	22.80
Livestock	11.86	26.37	18.48
Forestry	0.24	3.16	1.44
Fishing	0.48	5.57	1.05

Source: Authors' survey (2005, 2007, 2010, 2012).

Table B.17: Summary Statistics of Key Variables in Four Waves

First Wave: Year 2004 (N = 782)					
Variables	Obs	Mean	Std	Min	Max
Household income	782	6246	5128	0	50000
Agricultural income	782	3978	3862	0	37165
Nonagricultural income	782	2267	3239	0	50000
Household HH index of agricultural production (HHI [1]	758	0.46	0.16	0.19	1
Alternative measure (HHI [2])	758	0.46	0.16	0.19	1
Alternative measure (HHI [3])	758	0.46	0.16	0.19	1
Fertilizer use (yuan per mu)	716	152.63	106.20	1	1251
Hired labor cost (yuan per mu)	716	31.06	119.22	0.05	2001
Household size (migrants excluded)	782	3.69	1.55	0	8
Household land cultivated (mu)	781	3.66	2.76	0	20
Number of labor (age 16 - 60)	782	2.53	1.42	0	7
Highest education in the household (year)	780	5.43	3.31	0	14
Village leader in the household (dummy)	782	0.04	0.20	0	1
Second Wave: Year 2006 (N = 815)					
Variables	Obs	Mean	Std	Min	Max
Household income	815	7619	10413	0	223080
Agricultural income	815	3825	3822	0	33148
Nonagricultural income	815	3793	9666	0	223000
Household HH index of agricultural production (HHI [1]	765	0.41	0.16	0.18	1
Alternative measure (HHI [2])	765	0.41	0.16	0.18	1
Alternative measure (HHI [3])	765	0.41	0.15	0.18	1
Fertilizer use (yuan per mu)	722	255.53	251.97	1	3840
Hired labor cost (yuan per mu)	722	30.46	105.92	0.05	1932
Household size (migrants excluded)	815	3.35	1.66	0	10
Household land cultivated (mu)	810	3.90	2.99	0	20.5
Number of labor (age 16 - 60)	815	2.52	1.51	0	9
Highest education in the household (year)	811	6.12	3.47	0	18
Village leader in the household (dummy)	811	0.02	0.13	0	1
Third Wave: Year 2009 (N = 834)					
Variables	Obs	Mean	Std	Min	Max
Household income	834	11995	13934	0	191265
Agricultural income	834	5454	6614	0	67155
Nonagricultural income	834	6541	11869	0	182620
Household HH index of agricultural production (HHI [1]	759	0.49	0.18	0.22	1
Alternative measure (HHI [2])	759	0.47	0.18	0.19	1
Alternative measure (HHI [3])	759	0.47	0.18	0.21	1
Fertilizer use (yuan per mu)	700	244.93	199.24	1	2401
Hired labor cost (yuan per mu)	700	30.09	137.94	0.03	2144
Household size (migrants excluded)	834	3.85	1.83	0	12
Household land cultivated (mu)	806	3.10	2.92	0	32.5
Number of labor (age 16 - 60)	834	2.47	1.48	0	7
Highest education in the household (year)	833	6.17	3.59	0	18
Village leader in the household (dummy)	834	0.04	0.19	0	1
Fourth Wave: Year 2011 (N = 935)					
Variables	Obs	Mean	Std	Min	Max
Household income	935	16538	22166	1	250001
Agricultural income	935	5698	9623	0	109740
Nonagricultural income	935	10840	19455	1	250001
Household HH index of agricultural production (HHI [1]	763	0.47	0.19	0.20	1
Alternative measure (HHI [2])	763	0.47	0.19	0.20	1
Alternative measure (HHI [3])	763	0.48	0.19	0.20	1
Fertilizer use (yuan per mu)	665	399.87	1225.22	1	22576
Hired labor cost (yuan per mu)	665	26.59	232.59	1	4168
Household size (migrants excluded)	935	3.79	1.86	0	11
Household land cultivated (mu)	935	3.08	3.53	0	75
Number of labor (age 16 - 60)	935	2.49	1.69	0	8
Highest education in the household (year)	931	6.82	3.39	0	14
Village leader in the household (dummy)	851	0.03	0.17	0	1

Source : Authors' survey (2005, 2007, 2010, 2012).

Note : All the prices are deflated to year 2004.

Table B.18: Summary Statistics of Key Variables Between Villages by Road Access

Without Road Access (N = 610)					
Variables	Obs	Mean	Std	Min	Max
Household income	610	6644	7699	1	106614
Agricultural income	610	4139	6219	0	100513
Nonagricultural income	601	1484	3655	0	53951
Household HH index of agricultural production (HHI [1])	599	0.44	0.16	0.18	1
Alternative measure (HHI [2])	599	0.44	0.16	0.18	1
Alternative measure (HHI [3])	599	0.44	0.16	0.18	1
Fertilizer use (yuan per mu)	576	200.56	355.99	1	7469
Hired labor cost (yuan per mu)	576	10.00	105.87	0.05	2001
Household size (migrants excluded)	610	3.89	1.81	1	12
Household land cultivated (mu)	610	3.35	2.73	0	20.5
Number of labor (age 16 - 60)	610	2.62	1.52	0	9
Highest education in the household (year)	608	5.44	3.45	0	16
Village leader in the household (dummy)	604	0.05	0.21	0	1
With Road Access (N = 2533)					
Variables	Obs	Mean	Std	Min	Max
Household income	2533	12033	16555	1	250001
Agricultural income	2548	5213	6815	0	109740
Nonagricultural income	2531	5888	14064	0	250001
Household HH index of agricultural production (HHI [1])	2387	0.46	0.18	0.18	1
Alternative measure (HHI [2])	2387	0.46	0.17	0.18	1
Alternative measure (HHI [3])	2387	0.46	0.17	0.19	1
Fertilizer use (yuan per mu)	2227	276.43	678.34	1	22576
Hired labor cost (yuan per mu)	2227	34.67	165.21	0.03	4168
Household size (migrants excluded)	2548	3.72	1.71	0	12
Household land cultivated (mu)	2548	3.49	3.19	0	75
Number of labor (age 16 - 60)	2548	2.52	1.54	0	8
Highest education in the household (year)	2541	6.34	3.45	0	18
Village leader in the household (dummy)	2467	0.03	0.16	0	1

Source : Authors' survey (2005, 2007, 2010, 2012).

Note : All the prices are deflated to year 2004.

Table B.19: Impact of Road Access on Agricultural Production

	Dependent Variables: Agricultural specialization and input use (fertilizer and labor input)									
	HHI [1]		HHI [2]		HHI [3]		Fertilizer use (yuan per mu)		Hired labor cost (yuan per mu)	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Road*beforeafter	0.03***	0.03***	0.03***	0.03***	0.03**	0.02**	0.28***	0.29***	0.58***	0.56***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.10)	(0.08)	(0.15)	(0.15)
Clustered s.e. p-value:	0.00	0.01	0.00	0.01	0.01	0.04	0.01	0.01	0.00	0.00
Bootstrapped p-value:	0.00	0.01	0.01	0.02	0.03	0.11	0.02	0.01	0.01	0.02
Land		-0.00***		-0.00***		-0.00***		-0.07***		-0.03
		(0.00)		(0.00)		(0.00)		(0.01)		(0.03)
Number of primary age population (age 16 - 60)		0.00		0.00		0.00		0.04**		0.02
		(0.00)		(0.00)		(0.00)		(0.02)		(0.02)
Household size		-0.01***		-0.01***		-0.01***		0.04***		-0.05**
		(0.00)		(0.00)		(0.00)		(0.01)		(0.02)
Highest education (year)		0.00		0.00		0.00		0.00		0.00
		(0.00)		(0.00)		(0.00)		(0.01)		(0.01)
Village leader (dummy)		-0.04**		-0.04**		-0.04**		0.10		0.14
		(0.02)		(0.02)		(0.02)		(0.15)		(0.23)
Year fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Natural village fixed effec	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.07	0.10	0.06	0.08	0.06	0.09	0.04	0.07	0.13	0.13
AIC	-2392.8	-2456.3	-2442.8	-2496.2	-2467.9	-2526.1	8656.2	8563.2	11840.4	11837.5
N	2804	2804	2804	2804	2804	2804	2804	2804	2804	2804

Notes : 1. Households with no land are dropped from the regression for sample consistency through all the regressions. In addition, seven observations with self consumption ratio larger than one are dropped which are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at the natural village level. Bootstrapped p-values for the double difference coefficients obtained via wild bootstrap with Rademacher weights and imposing null hypothesis followed by Cameron, Gelbach, and Miller (2008).

Table B.20: Impact of Road Access on Income

	Dependent Variable: Agricultural income, nonfarm income, and income per capita					
	Agricultural income (log)		Nonfarm income (log)		Income per capita (log)	
	(1)	(2)	(1)	(2)	(1)	(2)
Road*beforeafter	0.17	0.24	0.14	0.23	0.1	0.09
	(0.13)	(0.14)	(0.32)	(0.31)	(0.11)	(0.11)
Clustered s.e. p-value:	0.22	0.12	0.66	0.46	0.38	0.43
Bootstrapped p-value:	0.26	0.15	0.69	0.49	0.41	0.46
Land		0.10***		0		0.06***
		(0.02)		(0.02)		(0.02)
Number of primary age population (age 16 - 60)		0.07***		0.19***		0.09***
		(0.01)		(0.04)		(0.01)
Household size		0.10***		0.18***		-0.16***
		(0.01)		(0.04)		(0.01)
Highest education (year)		0.00		0.04***		0.02***
		(0.01)		(0.01)		(0.00)
Village leader (dummy)		0.22*		0.34		0.28***
		(0.12)		(0.34)		(0.09)
Year fixed effect	YES	YES	YES	YES	YES	YES
Natural village fixed effect	YES	YES	YES	YES	YES	YES
R-squared	0.06	0.24	0.10	0.14	0.15	0.28
AIC	7808.9	7238.4	13455.5	13353.5	6951.5	6509.6
N	2804	2804	2804	2804	2804	2804

Notes : 1. Households with no land are dropped from the regression for sample consistency through all the regressions. In addition, seven observations with self consumption ratio larger than one are dropped which are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. When calculating the in-kind agricultural income, we use the market price of each agricultural product as our reference price. However, we also try to impose a 10 percent iceberg transportation cost on in-kind agricultural income to account for the transportation cost incurred during potential trading process. The result remains similar. 3. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 4. Robust standard errors are clustered at the natural village level. Bootstrapped p-values for the double difference coefficients obtained via wild bootstrap with Rademacher weights and imposing null hypothesis followed by Cameron, Gelbach, and Miller (2008).

Table B.21: Impact of Road Access on Agricultural Production (Placebo Test using Observations without Road Access)

	Dependent variables									
	HHI[1]		HHI[2]		HHI[3]		Fertilizer use (yuan per mu)		Hired labor cost (yuan per mu)	
Road*one period earlier	0.003	0.008	0.003	0.008	0.012	0.016	-0.143	-0.063	-0.302	-0.198
	(0.052)	(0.050)	(0.052)	(0.050)	(0.050)	(0.050)	(0.226)	(0.185)	(0.167)	(0.121)
N	576	576	576	576	576	576	576	576	576	576
Household characteristics	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Year fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Natural vllage fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes : 1. Short-run impact compares the villages without road until 2009 with the villages having road constructed in recent two years by the survey period; Long-run impact compares the villages without road until 2009 with villages having road since 2004; Difference between short-run and long-run impact compares villages having road since 2004 with villages having road constructed in recent two years. Due to space constraint, only the coefficients on "road*beforeafter" are reported. 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.

Table B.22: Impact of Road Access on Poverty Reduction (Aggregate Data at the Natural Village Level)

	Panel A: Poverty Measure—P 0			
	Low poverty line		*High poverty line*	
Road*beforeafter	-0.10***	-0.09***	-0.13***	-0.11*
	(0.03)	(0.03)	(0.04)	(0.06)
Administrative village fixed effect		YES		YES
R-squared	0.47	0.46	0.43	0.42
AIC	-140.6	-137.4	-83.7	-80.5
	Panel B: Poverty Measure—P 1			
	Low poverty line		*High poverty line*	
Road*beforeafter	-0.03***	-0.02*	-0.05***	-0.04**
	(0.01)	(0.02)	(0.01)	(0.01)
Administrative village fixed effect		YES		YES
R-squared	0.33	0.32	0.47	0.46
AIC	-262.4	-259.6	-222.3	-219.4
	Panel C: Poverty Measure—P 2			
	Low poverty line		*High poverty line*	
Road*beforeafter	-0.01**	-0.01	-0.03***	-0.02*
	(0.01)	(0.01)	(0.01)	(0.01)
Administrative village fixed effect		YES		YES
R-squared	0.17	0.15	0.35	0.34
AIC	-334.7	-331.3	-290.6	-287.6
N	68	68	68	68

Notes. 1. All the regressions use the aggregate data at the natural village level. Both low poverty line and high poverty line have been applied to calculate each poverty measure (P0, P1 and P2). Other control variables include 1) whether a natural village has road access by the end of year 2009; 2) acreage of land; 3) number of primary age (16-60) population; 4) household size; 5) dummy for village cadre and 6) year fixed effect; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 4. Robust standard errors are in the parentheses.

Table B.23: Robustness Checks on Area Based Specialization Index and Crop Yields

	Dependent Variable: Specialization index (area-based), maize yield, and rice yield					
	HHI (area-based)		Maize yield		Rice yield	
	(1)	(2)	(1)	(2)	(1)	(2)
Road*beforeafter	0.02	0.02	45.84*	44.91*	71.62	64.38
	(0.02)	(0.02)	(23.01)	(23.43)	(79.02)	(80.69)
Clustered s.e. p-value:	0.16	0.17	0.06	0.07	0.38	0.44
Bootstrapped p-value:	0.15	0.17	0.09	0.11	0.53	0.55
Land		-0.003**		-4.91*		-2.93
		(0.001)		(2.77)		(3.24)
Number of primary age population (age 16 - 60)		-0.005**		3.93		6.26**
		(0.002)		(3.05)		(2.46)
Household size		-0.006***		3.52		-1.23
		(0.002)		(2.21)		(4.85)
Highest education (year)		-0.00		-1.00		-1.66
		(0.00)		(1.39)		(1.84)
Village leader (dummy)		-0.02		29.34		13.97
		(0.02)		(23.80)		(47.13)
Year fixed effect	YES	YES	YES	YES	YES	YES
Natural village fixed effect	YES	YES	YES	YES	YES	YES
R-squared	0.05	0.07	0.19	0.20	0.10	0.10
AIC	-2039.28	-2061.49	17562.67	17551.79	7046.31	7053.10
N	2010	2010	1371	1371	541	541

Notes : 1. Households with no land are dropped from the regression for sample consistency through all the regressions. In addition, seven observations with self consumption ratio larger than one are dropped which are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. The observations in the year 2007 survey are dropped since no information on crop-specific cultivated areas is provided. 3. Only households with non-zero maize (rice) output are included in the regressions on maize (rice) yield. 4. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 5. Robust standard errors are clustered at the natural village level. Bootstrapped p-values for the double difference coefficients obtained via wild bootstrap with Rademacher weights and imposing null hypothesis followed by Cameron, Gelbach, and Miller (2008).

Table B.24: Summary of Statistics about Home Demolition from Household Survey

	All	Eastern China	Central & Western China
Share of households experienced demolition	8.70%	11.54%***	5.03%
Average square meters being demolished	129.15	136.52**	107.66
Share of households without any compensation	4.73%	2.84%**	10.38%
Average monetary compensation (10,000 yuan)	27.68	33.30***	10.15
Share of households not satisfied with the compensation	47.26%	47.88%	45.26%
Total number of households	4,875	2,748	2,127

Data source: China Household Finance Survey (CHFS) 2011.

Notes: 1. Only urban households are included in this table. 2. Households with the most recent demolition earlier than year 2000 are excluded from this table. 3. According to the definition of CHFS, eastern China in the surveyed sample includes Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Shandong and Guangdong; central China includes Shanxi, Anhui, Jiangxi, Henan, Hubei and Hunan; western China include Jilin, Heilongjiang, Chongqing, Sichuan, Yunnan, Gansu and Qinghai. 4. The t-test results between groups are reported. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table B.25: Summary of Statistics

Variable	Source	Unit of Measurement	Mean	Standard Deviation	Minimum	Maximum
Household Savings Rate	*CEIC*	% of savings rate	28.30	6.76	3.91	50.86
Forced Eviction Index	*Baidu News Archive*	number of news reported	7.54	44.47	0	1280
GDP	*China Economic and Social Development Statistical Database*	100 million *yuan*	1055.52	1252.19	48.23	12423.44
Fixed Asset Investment		100 million *yuan*	37.50	772.00	0.21	26000
Total Population		10 thousand	430.24	237.48	45.91	1270.19
Number of Cities = 248; Number of Years=8						

Table B.26: Factors Affecting Forced Evictions

	Dependent Variable: Ln Forced Eviction Index				
	[1]	[2]	[3]	[4]	[5]
Ln GDP (lagged)	0.43***				0.46***
	(0.06)				(0.10)
Ln Fixed Asset Investment (lagged)		0.24**			0.02
		(0.10)			(0.03)
Ln Total Population (lagged)			0.21***		-0.05
			(0.05)		(0.07)
Ln Foreign Direct Investment (lagged)				0.13***	0.00
				(0.03)	(0.03)
Provincial Capital = 1	1.59***	1.82***	2.04***	1.87***	1.56***
	(0.16)	(0.22)	(0.16)	(0.17)	(0.17)
Municipality = 1	1.81***	2.28***	2.80***	2.41***	1.69***
	(0.18)	(0.24)	(0.14)	(0.14)	(0.24)
Year Fixed Effects	Y	Y	Y	Y	Y
Province Fixed Effects	Y	Y	Y	Y	Y
R-Squared	0.57	0.54	0.54	0.55	0.57
N	1,701	1,589	1,701	1,695	1,583

Notes: 1. Robust tandard errors clustered at the city level are reported in parentheses. 2. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table B.27: The Impact of Reported Forced Evictions on Household Savings Rate (Difference GMM Estimator)

	Panel A: Difference GMM with all the RHS variables assumed to be exogenous				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.39**	0.39**	0.41***	0.38**	0.39**
	(0.19)	(0.18)	(0.18)	(0.18)	(0.16)
Number of instruments	17	17	17	17	20
Sargan Test (p-value)	18.02 (0.12)	16.91 (0.15)	17.04 (0.15)	14.84 (0.25)	15.34 (0.22)
First order serial correlation test (p value)	-5.91 (0.00)	-5.71 (0.00)	-5.85 (0.00)	-5.93 (0.00)	-5.84 (0.00)
Second order serial correlation test (p value)	-1.12 (0.26)	-0.99 (0.32)	-1.19 (0.24)	-1.25 (0.21)	-0.96 (0.34)
	Panel B: Difference GMM with all the RHS variables assumed to be endogenous				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.67**	0.88*	0.29	0.68***	0.37
	(0.28)	(0.45)	(0.26)	(0.25)	(0.29)
Number of instruments	35	35	35	35	50
Sargan Test (p-value)	32.69 (0.34)	28.64 (0.54)	29.80 (0.48)	30.15 (0.46)	9.35 (0.67)
First order serial correlation test (p value)	-6.10 (0.00)	-5.25 (0.00)	-5.99 (0.00)	-6.12 (0.00)	-5.69 (0.00)
Second order serial correlation test (p value)	-1.09 (0.28)	-0.66 (0.51)	-1.06 (0.29)	-1.17 (0.24)	-0.96 (0.34)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Number of Observations	1,177	1,142	1,177	1,167	1,132

Notes: 1. All regressions are estimated using the Difference GMM Estimator proposed by Arellano and Bond (1991). 2. Robust standard errors are reported in parentheses. 3. *** significant at 1%, ** significant at 5%, * significant at 10%. 3. In addition to the four control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications. A maximum of three lagged dependent variables can be used as instruments for the differenced lag dependent variable. 4. For Panel B, all the second and higher order lagged values of forced eviction index are used as instruments; the second order lagged values of economic control variables are used as instruments.

Table B.28: The Impact of Reported Forced Evictions on Household Savings Rate (Difference GMM Estimator, Alternative Measure)

	Panel A: Differenced GMM with all the RHS variables assumed to be exogenous				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.31*	0.29	0.30	0.29	0.26
	(0.18)	(0.18)	(0.19)	(0.18)	(0.19)
Number of instruments	17	17	17	17	20
Sargan Test (p-value)	18.24 (0.11)	18.10 (0.11)	18.35 (0.11)	15.78 (0.20)	15.68 (0.21)
First order serial correlation test (p value)	-5.94 (0.00)	-5.73 (0.00)	-5.89 (0.00)	-5.97 (0.00)	-5.87 (0.00)
Second order serial correlation test (p value)	-1.06 (0.29)	-0.94 (0.35)	-1.11 (0.27)	-1.20 (0.23)	-0.93 (0.35)
	Panel B: Differenced GMM with all the RHS variables assumed to be endogenous				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.90**	0.59*	0.06	0.93**	0.34
	(0.38)	(0.35)	(0.43)	(0.37)	(0.39)
Number of instruments	35	35	35	35	50
Sargan Test (p-value)	32.77 (0.33)	31.86 (0.37)	27.66 (0.59)	29.15 (0.51)	40.79 (0.52)
First order serial correlation test (p value)	-6.07 (0.00)	-5.28 (0.00)	-6.05 (0.00)	-6.12 (0.00)	-5.53 (0.00)
Second order serial correlation test (p value)	-0.96 (0.34)	-0.57 (0.57)	-0.93 (0.35)	-1.07 (0.28)	-0.92 (0.36)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Number of Observations	1,177	1,142	1,177	1,167	1,132

Notes: 1. All regressions are estimated using the Difference GMM Estimator proposed by Arellano and Bond (1991). 2. Robust standard errors are reported in parentheses. 3. *** significant at 1%, ** significant at 5%, * significant at 10%. 3. In addition to the four control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications. A maximum of three lagged dependent variables can be used as instruments for the differenced lag dependent variable. 4. For Panel B, all the second and higher order lagged values of forced eviction index are used as instruments; the second order lagged values of economic control variables are used as instruments.

Table B.29: The Impact of Reported Forced Evictions on Household Savings Rate (Difference GMM Estimator, Normalized Measure)

	Panel A: Differenced GMM with all the RHS variables assumed to be exogenous				
	[2]	[3]	[4]	[5]	[6]
Ln Forced Eviction Index (lagged)	291.44***	271.55***	264.18***	277.71***	299.33***
	(82.09)	(76.95)	(79.10)	(83.87)	(96.23)
Number of instruments	17	17	17	17	20
Sargan Test (p-value)	18.77 (0.09)	18.53 (0.10)	18.70 (0.10)	15.47 (0.22)	15.98 (0.19)
First order serial correlation test (p value)	-5.99 (0.00)	-5.79 (0.00)	-5.93 (0.00)	-5.97 (0.00)	-5.86 (0.00)
Second order serial correlation test (p value)	-1.16 (0.24)	-1.04 (0.30)	-1.21 (0.23)	-1.32 (0.19)	-1.05 (0.29)
	Panel B: Differenced GMM with all the RHS variables assumed to be endogenous				
	[2]	[3]	[4]	[5]	[6]
Ln Forced Eviction Index (lagged)	236.36	173.86	195.52	294.85	127.73
	(254.57)	(255.08)	(186.80)	(259.70)	(199.63)
Number of instruments	35	35	35	35	50
Sargan Test (p-value)	35.54 (0.22)	31.93 (0.37)	34.43 (0.26)	33.64 (0.30)	51.71 (0.14)
First order serial correlation test (p value)	-6.34 (0.00)	-5.44 (0.00)	-6.14 (0.00)	-6.50 (0.00)	-6.04 (0.00)
Second order serial correlation test (p value)	-1.04 (0.30)	-0.67 (0.50)	-1.40 (0.16)	-1.25 (0.21)	-0.96 (0.34)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Number of Observations	1,177	1,142	1,177	1,167	1,132

Notes: 1. All regressions are estimated using the Difference GMM Estimator proposed by Arellano and Bond (1991). 2. Robust standard errors are reported in parentheses. 3. *** significant at 1%, ** significant at 5%, * significant at 10%. 3. In addition to the four control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications. A maximum of three lagged dependent variables can be used as instruments for the differenced lag dependent variable. 4. For Panel B, all the second and higher order lagged values of forced eviction index are used as instruments; the second order lagged values of economic control variables are used as instruments.

Table B.30: The Impact of Forced Eviction on Home Sale (Differenced GMM)

	Dependent variable: residential floor area sold (thousand square meters)				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	-0.12***	-0.07**	0.06*	0.00	-0.09***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Ln Forced Eviction Index (alternative, lagged)	-0.15***	-0.12***	-0.02	-0.07**	-0.11***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Ln Forced Eviction Index (normalized)	-30.97	-20.80	-11.65	-9.81	-17.39
	(22.72)	(16.31)	(11.32)	(11.40)	(17.74)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Number of Observations	992	957	992	984	949

Notes: 1. All regressions are estimated using the Difference GMM Estimator proposed by Arellano and Bond (1991). All the RHS variables are specified as endogenous variables. all the second and higher order lagged values of forced eviction index are used as instruments; the second order lagged values of economic control variables are used as instruments. 2. Due to space constraint, only the coefficients on three measures of forced eviction are reported. The specification tests (Sargan test and serial correlation test) support the validity of the models. 3. Robust standard errors are reported in parentheses. 4. *** significant at 1%, ** significant at 5%, * significant at 10%. 5. In addition to the four control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications. A maximum of three lagged dependent variables can be used as instruments for the differenced lag dependent variable.

Table B.31: The Impact of Reported Forced Evictions on Household Savings Rate (Difference GMM Estimator, Restricted Sample)

	Panel A: Restricted sample: Central&Western China				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.89**	0.72*	0.62	0.91***	0.89**
	(0.38)	(0.48)	(0.45)	(0.36)	(0.44)
Ln Forced Eviction Index (alternative, lagged)	1.23*	1.02*	0.79	1.38**	1.12*
	(0.63)	(0.60)	(0.62)	(0.59)	(0.66)
Ln Forced Eviction Index (normalized)	1074.83**	1033.45**	568.22	1268.05**	900.61
	(534.62)	(494.65)	(648.00)	(539.22)	(708.34)
Number of Observations	671	637	671	663	629
	Panel B: Restricted sample: Excluding provincial capitals				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.85**	0.41	0.71**	0.80**	0.49
	(0.39)	(0.39)	(0.33)	(0.34)	(0.21)
Ln Forced Eviction Index (alternative, lagged)	1.42**	0.73	1.21*	1.35**	0.77
	(0.58)	(0.54)	(0.63)	(0.55)	(0.60)
Ln Forced Eviction Index (normalized)	-195.72	-244.28	-219.39	-175.43	-326.19
	(207.00)	(183.37)	(221.17)	(187.32)	(152.03)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Number of Observations	1,067	1,032	1,067	1,057	1,022

Notes: 1. All regressions are estimated using the Difference GMM Estimator proposed by Arellano and Bond (1991). All the RHS variables are specified as endogenous variables. all the second and higher order lagged values of forced eviction index are used as instruments; the second order lagged values of economic control variables are used as instruments. 2. Due to space constraint, only the coefficients on three measures of forced eviction are reported. The specification tests (Sargan test and serial correlation test) support the validity of the models. 3. Robust standard errors are reported in parentheses. 4. *** significant at 1%, ** significant at 5%, * significant at 10%. 5. In addition to the four control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications. A maximum of three lagged dependent variables can be used as instruments for the differenced lag dependent variable.

Table B.32: The Impact of Reported Forced Evictions on Household Savings Rate (System GMM Estimator)

	Specification: System GMM with all the RHS variables assumed to be endogenous				
	[1]	[2]	[3]	[4]	[5]
Ln Forced Eviction Index (lagged)	0.46	0.35	-1.00	0.22	0.12
	(0.61)	(0.43)	(0.78)	(0.50)	(0.60)
Number of instruments	22	22	22	22	37
Sargan Test (p-value)	14.67 (0.26)	10.91 (0.54)	16.35 (0.18)	17.42 (0.13)	44.29 (0.01)
First order serial correlation test (p value)	-0.96 (0.34)	-0.89 (0.37)	-1.03 (0.30)	-0.32 (0.75)	-1.06 (0.29)
Second order serial correlation test (p value)	-0.54 (0.59)	-1.09 (0.27)	-0.71 (0.48)	-1.80 (0.07)	-2.07 (0.04)
Ln Forced Eviction Index (alternative, lagged)	0.67	0.53	-0.83	0.26	0.50
	(0.67)	(0.57)	(0.78)	(0.64)	(0.70)
Number of instruments	22	22	22	22	37
Sargan Test (p-value)	13.86 (0.31)	9.56 (0.66)	20.59 (0.06)	13.69 (0.32)	42.57 (0.01)
First order serial correlation test (p value)	-0.78 (0.44)	-0.76 (0.45)	-0.92 (0.36)	-0.21 (0.83)	-1.04 (0.30)
Second order serial correlation test (p value)	-0.66 (0.51)	-1.17 (0.24)	-0.73 (0.46)	-1.87 (0.06)	-2.14 (0.03)
Ln Forced Eviction Index (normalized)	-262.60	-189.47	-58.47	-121.72	-490.09**
	(300.09)	(283.59)	(238.47)	(458.05)	(198.12)
Number of instruments	22	22	22	22	37
Sargan Test (p-value)	19.32 (0.08)	13.01 (0.37)	17.97 (0.12)	13.73 (0.32)	45.51 (0.01)
First order serial correlation test (p value)	-0.96 (0.34)	-0.96 (0.34)	-0.88 (0.38)	-0.33 (0.74)	-0.97 (0.33)
Second order serial correlation test (p value)	-0.65 (0.51)	-1.07 (0.28)	-0.80 (0.42)	-1.52 (0.13)	-2.83 (0.01)
Lagged Control Variables					
Ln GDP (lagged)	Y				Y
Ln Fixed Asset Investment (lagged)		Y			Y
Ln Total Population (lagged)			Y		Y
Ln Foreign Direct Investment (lagged)				Y	Y
Year Dummies	Y	Y	Y	Y	Y
Number of Observations	1,421	1,403	1,421	1,415	1,397

Notes: 1. All regressions are estimated using the System GMM Estimator proposed by Blundell and Bond (1998). 2. All the RHS variables are treated as endogenous variables (except for year dummies), using lags 2 and deeper as instruments for the transformed equation and lag 1 for the levels equation, which is the standard treatment for endogenous variables (Roodman, 2006). Collapsed instrument sets have been applied in all the specifications to avoid the "too many instruments" problem. 3. Robust standard errors are reported in parentheses. 4. *** significant at 1%, ** significant at 5%, * significant at 10%. 5. In addition to the control variables mentioned in the table, the first and second order lagged savings rate are also covariates for all the specifications which is consistent with the difference GMM models.

BIBLIOGRAPHY

[1] Gabriel M. Ahlfeldt. The train has left the station: Do markets value intracity access to intercity rail connections? *German Economic Review*, 12(3):312–335, 08 2011.

[2] D. Albalate and G. Bel. *The Economics and Politics of High-speed Rail: Lessons from Experiences Abroad*. Lexington Books, 2012.

[3] Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

[4] Manuel Arellano and Olympia Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

[5] Nikos Askitas and Klaus F Zimmermann. Google econometrics and unemployment forecasting. 2009.

[6] Jeremy Atack, Fred Bateman, Michael Haines, and Robert A Margo. Did railroads induce or follow economic growth? urbanization and population growth in the american midwest, 1850–1860. *Social Science History*, 34(2):171–197, 2010.

[7] Chun-Chung Au and J Vernon Henderson. Are chinese cities too small? *The Review of Economic Studies*, 73(3):549–576, 2006.

[8] Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in china. 2009. MIT Working Paper.

[9] Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in china. 2012. NBER Working Paper No. 17897.

[10] Nathaniel Baum-Snow. Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805, 2007.

[11] Nathaniel Baum-Snow, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang. Roads, railroads and decentralization of chinese cities. 2012. Working Paper.

[12] Vincent Benziger. Urban access and rural productivity growth in post-mao china. *Economic Development and Cultural Change*, 44(3):539–70, 1996.

[13] Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.

[14] Olivier Blanchard and Francesco Giavazzi. Rebalancing growth in china: A three-handed approach. *China & World Economy*, 14(4):1–20, 2006.

[15] Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143, 1998.

[16] Richard Bullock, Andrew Salzberg, and Ying Jin. High-speed rail–the first three years: Taking the pulse of china's emerging program. 2012. China Transport Topics No. 04, World Bank.

[17] César Calderón. Infrastructure and growth in africa. 2009.

[18] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

[19] John Y Campbell and Joao F Cocco. How do house prices affect consumption? evidence from micro data. *Journal of Monetary Economics*, 54(3):591–621, 2007.

[20] Marcos D Chamon and Eswar S Prasad. Why are saving rates of urban households in china rising? *American Economic Journal: Macroeconomics*, pages 93–130, 2010.

[21] Hyunyoung Choi and Hal Varian. Predicting the present with google trends. *Economic Record*, 88(s1):2–9, 2012.

[22] Saugato Datta. The impact of improved highways on indian firms. *Journal of Development Economics*, 99(1):46–57, 2012.

[23] Dave Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 2013. forthcoming.

[24] Dave Donaldson and Richard Hornbeck. Railroads and american economic growth:a "market access" approach. 2013. NBER Working Paper No. 19213.

[25] Paul Dorosh, Hyoung-Gun Wang, Liang You, and Emily Schmidt. Crop production and road connectivity in sub-saharan africa: a spatial analysis. 2010.

[26] Esther Duflo and Rohini Pande. Dams. *The Quarterly Journal of Economics*, 122(2):601–646, 2007.

[27] Economist. The great train robbery. 9 2011. http://www.economist.com/node/21528263.

[28] Gary V Engelhardt. House prices and home owner saving behavior. *Regional Science and Urban Economics*, 26(3):313–336, 1996.

[29] Javier Escobal and Carmen Ponce. The benefits of rural roads: Enhancing income opportunities for the rural poor. *Documento de trabajo*, (40), 2002.

[30] Benjamin Faber. Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. 2013. Working Paper.

[31] Shenggen Fan and Connie Chan-Kang. *Road Development, Economic Growth, and Poverty Reduction in China*. Number 12. International Food Policy Research Institute, 2005.

[32] Shenggen Fan and Peter Hazell. Returns to public investments in the less-favored areas of india and china. *American Journal of Agricultural Economics*, 83(5):1217–1222, 2001.

[33] Shenggen Fan and Xiaobo Zhang. Infrastructure and regional economic development in rural china. *China Economic Review*, 15(2):203–214, 2004.

[34] Jin Feng, Lixin He, and Hiroshi Sato. Public pension and household saving: Evidence from urban china. *Journal of Comparative Economics*, 39(4):470–485, 2011.

[35] James Foster, Joel Greer, and Erik Thorbecke. A class of decomposable poverty measures. *Econometrica: Journal of the Econometric Society*, pages 761–766, 1984.

[36] Masahisa Fujita, Paul Krugman, and Anthony J. Venables. *The Spatial Economy*. MIT press, 2001.

[37] Ejaz Ghani, Arti Grover Goswami, and William R Kerr. Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. Technical report, 2012. NBER Working Paper No. 18524.

[38] John Gibson and Scott Rozelle. Poverty and access to roads in papua new guinea. *Economic Development and Cultural Change*, 52(1):159–185, 2003.

[39] John Giles and Kyeongwon Yoo. Precautionary behavior, migrant networks, and household consumption decisions: An empirical analysis using household panel data from rural china. *The Review of Economics and Statistics*, 89(3):534–551, 2007.

[40] Douglas Gollin and Richard Rogerson. Agriculture, roads, and economic development in uganda. Technical report, National Bureau of Economic Research, 2010.

[41] Kaddour Hadri. Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3(2):148–161, 2000.

[42] Richard DF Harris and Elias Tzavalis. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91(2):201–226, 1999.

[43] Xiaoqing Hu. Portfolio choices for homeowners. *Journal of Urban Economics*, 58(1):114–136, 2005.

[44] International Union of Railways (UIC). High speed lines in the world. 2010. http://www.uic.org/IMG/pdf/20101219_d_high_speed_lines_in_the_world_maps.pdf.

[45] Hanan G Jacoby. Access to markets and the benefits of rural roads. *The Economic Journal*, 110(465):713–737, 2000.

[46] Hanan G Jacoby and Bart Minten. On measuring the benefits of lower transport costs. *Journal of Development Economics*, 89(1):28–38, 2009.

[47] Jyotsna Jalan and Martin Ravallion. Geographic poverty traps? a micro model of consumption growth in rural china. *Journal of Applied Econometrics*, 17(4):329–346, 2002.

[48] Ravi Kanbur and Xiaobo Zhang. Fifty years of regional inequality in china: a journey through central planning, reform, and openness. *Review of Development Economics*, 9(1):87–106, 2005.

[49] Shahidur R Khandker, Zaid Bakht, and Gayatri B Koolwal. The poverty impact of rural roads: Evidence from bangladesh. 2006.

[50] Vasco Leite, Sofia BSD Castro, and João Correia-da Silva. A third sector in the core-periphery model: Non-tradable goods. *The Annals of Regional Science*, 50(1):71–108, 2013.

[51] Andrew Levin, Chien-Fu Lin, and Chia-Shang James Chu. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24, 2002.

[52] Nuno Limao and Anthony J Venables. Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review*, 15(3):451–479, 2001.

[53] Juan Liu. tie lu shi shi kua yue shi fa zhan zhan lue. 2003. http://www.curb.com.cn/pageshow.asp?id_forum=000106.

[54] Xin Meng. Unemployment, consumption smoothing, and precautionary saving in urban china. *Journal of Comparative Economics*, 31(3):465–485, 2003.

[55] Guy Michaels. The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701, 2008.

[56] Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426, 1981.

[57] Steven Were Omamo. Farm-to-market transaction costs and specialisation in small-scale agriculture: Explorations with a non-separable household model. *The Journal of Development Studies*, 35(2):152–163, 1998.

[58] Duo Qin, Marie Anne Cagas, Pilipinas Quising, and Xin-Hua He. How much does investment drive economic growth in china? *Journal of Policy Modeling*, 28(7):751–774, 2006.

[59] Debraj Ray. Uneven growth: A framework for research in development economics. *The Journal of Economic Perspectives*, 24(3):45–60, 2010.

[60] Mitch Renkow, Daniel G Hallstrom, and Daniel D Karanja. Rural infrastructure, transactions costs and market participation in kenya. *Journal of Development Economics*, 73(1):349–367, 2004.

[61] Sina. ti su shi cheng shi jian huo che lv xing shi jian zong ti ya suo 20%-30%, 2007. http://news.sina.com.cn/c/2007-04-12/151112762996.shtml.

[62] State Council. zhong chang qi tie lu wang gui hua, 2004. http://jtyss.ndrc.gov.cn/fzgh/t20050720_37500.htm.

[63] State Council. zhong chang qi tie lu wang gui hua, (2008 nian tiao zheng), 2008. http://jtyss.ndrc.gov.cn/fzgh/W020090605632915547512.pdf.

[64] David Stifel, Bart Minten, and Paul Dorosh. Transaction costs and agricultural productivity: Implications of isolation for rural poverty in madagascar. *Markets and Structural Studies Division (MSSD) Discussion Paper*, (56), 2003.

[65] Simeon Vosen and Torsten Schmidt. Forecasting private consumption: Survey-based indicators vs. google trends. *Journal of Forecasting*, 30(6):565–578, 2011.

[66] Wall Street Journal. China ndrc: Speeding up infrastructure construction this year. 2013. http://online.wsj.com/article/BT-CO-20130804-701739.html.

[67] Shang-Jin Wei and Xiaobo Zhang. The competitive saving motive: Evidence from rising sex ratios and savings rates in china. *Journal of Political Economy*, 119(3):511–564, 2011.

[68] Li Xing, Shenggen Fan, Xiaopeng Luo, and Xiaobo Zhang. Community poverty and inequality in western china: A tale of three villages in guizhou province. *China Economic Review*, 20(2):338–349, 2009.

[69] Chenggang Xu. The fundamental institutions of china's reforms and development. *Journal of Economic Literature*, 49(4):1076–1151, 2011.

[70] Xiaokai Yang and Yew-Kwang Ng. *Specialization and Economic Organization: A New Classical Microeconomic Framework*. North-Holland Amsterdam, 1993.

[71] Qiao Yu. Capital investment, international trade and economic growth in china: Evidence in the 1980–1990s. *China Economic Review*, 9(1):73–84, 1998.

[72] Xiaobo Zhang and Shenggen Fan. How productive is infrastructure? a new approach and evidence from rural india. *American Journal of Agricultural Economics*, 86(2):492–501, 2004.

[73] Xiaobo Zhang, Jin Yang, and Shenglin Wang. China has reached the lewis turning point. *China Economic Review*, 22(4):542–554, 2011.

[74] Siqi Zheng and Matthew E Kahn. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences*, 110(14):E1248–E1253, 2013.